

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:51:24 ; Search time 32.998 Seconds
(without alignments)
3180.293 Million cell updates/sec

itle: US-09-941-947A-24

irect score: 2598

equences: 1 MNSNDNRQVIVIGAGLGGLS.....MPMVTSCQLVRDKIVADLQ 497

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 809742 seqs, 211153259 residues

otal number of hits satisfying chosen parameters: 809742

inimum DB seq length: 0

aximum DB seq length: 2000000000

et-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2598	100.0	497	9	US-09-934-903-18
2	2598	100.0	497	9	US-09-934-868-78
3	2598	100.0	497	10	US-09-941-947A-24
4	2598	100.0	497	14	US-10-358-917-8
5	1289	49.6	497	14	US-10-358-917-14
6	1039	40.0	439	9	US-09-815-242-5696
7	809	31.1	517	15	US-10-369-493-19519
8	687	26.4	492	10	US-09-941-947A-32
9	687	26.4	492	14	US-10-218-118-8
10	674	25.9	491	15	US-10-369-493-234
11	672	25.9	502	14	US-10-358-917-12
12	662.5	25.5	512	15	US-10-369-493-18644
13	654	25.2	499	15	US-10-369-493-18983
14	645	24.8	498	15	US-10-369-493-20438
15	637	24.5	494	9	US-09-547-267-5

16	637	24.5	494	10	US-09-920-923-4	Sequence 4, Appli
17	628	24.2	494	15	US-10-369-493-8951	Sequence 8951, Ap
18	619.5	23.8	498	15	US-10-369-493-19248	Sequence 19248, A
19	617	23.7	506	15	US-10-369-493-17830	Sequence 17830, A
20	607.5	23.4	511	9	US-09-934-903-16	Sequence 16, Appl
21	607.5	23.4	511	9	US-09-934-888-76	Sequence 76, Appl
22	607.5	23.4	511	10	US-09-941-947A-22	Sequence 22, Appl
23	607.5	23.4	511	14	US-10-358-917-4	Sequence 4, Appli
24	596	22.9	290	9	US-09-939-980-462	Sequence 462, Appl
25	590.5	22.7	514	15	US-10-369-493-21541	Sequence 21541, A
26	586	22.6	618	15	US-10-369-493-3769	Sequence 3769, Ap
27	576	22.2	498	15	US-10-369-493-10579	Sequence 10579, A
28	568.5	21.9	536	15	US-10-369-493-18646	Sequence 18646, A
29	564.5	21.7	496	15	US-10-369-493-20440	Sequence 20440, A
30	551	21.2	518	15	US-10-369-493-7750	Sequence 7750, Ap
31	550.5	21.2	495	15	US-10-369-493-7895	Sequence 7895, Ap
32	544	20.9	548	15	US-10-369-493-540	Sequence 540, App
33	536	20.6	506	15	US-10-369-493-8142	Sequence 8142, Ap
34	529.5	20.4	513	14	US-10-156-761-8562	Sequence 8562, Ap
35	522.5	20.1	1268	15	US-10-438-784-3	Sequence 3, Appli
36	499.5	19.2	498	15	US-10-369-493-10196	Sequence 10196, A
37	475	18.3	548	9	US-09-738-626-4194	Sequence 4194, Ap
38	469.5	18.1	469	15	US-10-438-784-5	Sequence 5, Appli
39	458	17.6	485	15	US-10-369-493-19518	Sequence 19518, A
40	440.5	17.0	404	15	US-10-369-493-19547	Sequence 19547, A
41	440	16.9	497	14	US-10-156-761-13629	Sequence 13629, A
42	411	15.8	530	14	US-10-128-713A-18	Sequence 18, Appl
43	386	14.9	498	15	US-10-369-493-17259	Sequence 17259, A
44	357	13.7	492	15	US-10-369-493-19911	Sequence 19911, A
45	347	13.4	494	15	US-10-369-493-18944	Sequence 18944, A

ALIGNMENTS

RESULT 1

```
US-09-934-903-18
; Sequence 18, Application US/09934903
; Patent No. US20020102690A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, Mattheos
; APPLICANT: Odum, J. Martin
; APPLICANT: Schenzle, Andreas J.
; APPLICANT: No. US20020102690Alton, Kelley C.
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Rouviere, Pierre
; APPLICANT: Picataggio, Stephen
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
; FILE REFERENCE: CL1646 US NA
; CURRENT APPLICATION NUMBER: US/09/934,903
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: September 1, 2001
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ORF9
US-09-934-903-18
```

Query Match 100.0%; Score 2598; DB 9; Length 497;

Best Local Similarity 100.0%; Pred. No. 4.6e-251;

Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNSNDNRQVIVIGAGLGGLSAISLATAGFVSQVLEKNDKVGKLNIMTKGFTFDLQPS 60

Db 1 MNSNDNRQVIVIGAGLGGLSAISLATAGFVSQVLEKNDKVGKLNIMTKGFTFDLQPS 60

Qy 61 ILTMPIFEALFTGAGKNMADYVQIQKVEPHWRNPFEDGSVIDLCEDAETQRRELDKLP 120

Db 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLP 120
Qy 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFGPLRLSLLSFDVFRSMDQGVRR 180
Db 121 GTYAQORFLDYSKNLTETAGYFAKLDGFDNLLKFGPLRLSLLSFDVFRSMDQGVRR 180
Qy 181 FISDPKLVILNYFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240
Db 181 FISDPKLVILNYFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240
Qy 241 ELGVIRLDAEVSLEIKQDGRACAVKLANGDVLPAIDIVSNMEVI PAMEKLLRSPASELK 300
Db 241 ELGVIRLDAEVSLEIKQDGRACAVKLANGDVLPAIDIVSNMEVI PAMEKLLRSPASELK 300
Qy 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVAP 360
Db 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVAP 360
Qy 361 CKTDPAPAGCEIIKILPHLDPKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420
Db 361 CKTDPAPAGCEIIKILPHLDPKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420
Qy 421 EBYWTFDIOQAKYYSNOGSIYGVVADRFKNLGFAPQRSSELSNLYFVGGSVNPGGMPM 480
Db 421 EBYWTFDIOQAKYYSNOGSIYGVVADRFKNLGFAPQRSSELSNLYFVGGSVNPGGMPM 480
Qy 481 VTLSGQLVRDKIVADLQ 497
Db 481 VTLSGQLVRDKIVADLQ 497

RESULT 2

US-09-934-868-78
; Sequence 78, Application US/09934868
; Patent No. US20020137190A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, James M
; APPLICANT: Odom, James J
; APPLICANT: Schenzle, Andreas J
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 78
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by CRTW2
US-09-934-868-78

Query Match 100.0%; Score 2598; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 4.6e-251;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGCKLNIMTKDGTFFDLGSP 60
Db 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGCKLNIMTKDGTFFDLGSP 60
Qy 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLP 120
Db 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLP 120
Qy 121 GTYAQORFLDYSKNLTETAGYFAKLDGFDNLLKFGPLRLSLLSFDVFRSMDQGVRR 180
Db 121 GTYAQORFLDYSKNLTETAGYFAKLDGFDNLLKFGPLRLSLLSFDVFRSMDQGVRR 180
Qy 181 FISDPKLVILNYFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240

Db 181 FISDPKLVILNYFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240
Qy 241 ELGVIRLDAEVSLEIKQDGRACAVKLANGDVLPAIDIVSNMEVI PAMEKLLRSPASELK 300
Db 241 ELGVIRLDAEVSLEIKQDGRACAVKLANGDVLPAIDIVSNMEVI PAMEKLLRSPASELK 300
Qy 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVAP 360
Db 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVAP 360
Qy 361 CKTDPAPAGCEIIKILPHLDPKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420
Db 361 CKTDPAPAGCEIIKILPHLDPKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420
Qy 421 EBYWTFDIOQAKYYSNOGSIYGVVADRFKNLGFAPQRSSELSNLYFVGGSVNPGGMPM 480
Db 421 EBYWTFDIOQAKYYSNOGSIYGVVADRFKNLGFAPQRSSELSNLYFVGGSVNPGGMPM 480
Qy 481 VTLSGQLVRDKIVADLQ 497
Db 481 VTLSGQLVRDKIVADLQ 497

RESULT 3

US-09-941-947A-24
; Sequence 24, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Methylobionas 16a
US-09-941-947A-24

Query Match 100.0%; Score 2598; DB 10; Length 497;
Best Local Similarity 100.0%; Pred. No. 4.6e-251;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGCKLNIMTKDGTFFDLGSP 60
Db 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGCKLNIMTKDGTFFDLGSP 60
Qy 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLP 120
Db 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLP 120
Qy 121 GTYAQORFLDYSKNLTETAGYFAKLDGFDNLLKFGPLRLSLLSFDVFRSMDQGVRR 180
Db 121 GTYAQORFLDYSKNLTETAGYFAKLDGFDNLLKFGPLRLSLLSFDVFRSMDQGVRR 180
Qy 181 FISDPKLVILNYFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240
Db 181 FISDPKLVILNYFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240

241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVVSNMEVIPAMEKLRSPASELK 300
|||||
241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVVSNMEVIPAMEKLRSPASELK 300
|||||
301 KXQRFEPSCGLVHLGLVDRLYPQLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
|||||
301 KXQRFEPSCGLVHLGLVDRLYPQLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
|||||
361 CKTDPAQAPACEIILPHILPDLDPKLLTAEDYSALRERVLVKLRMGLTDLRQHIVT 420
|||||
361 CKTDPAQAPACEIILPHILPDLDPKLLTAEDYSALRERVLVKLRMGLTDLRQHIVT 420
|||||
421 EEWTPDLIDQAKYYSNQSISYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480
|||||
421 EEWTPDLIDQAKYYSNQSISYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480
|||||
481 VTLGQVLVDKIVADLQ 497
|||||
481 VTLGQVLVDKIVADLQ 497
|||||

SULT 4
-10-358-917-8
Sequence 5, Application US/10358917
Publication No. US20030182687A1
GENERAL INFORMATION:
APPLICANT: Cheng, Qiong
APPLICANT: No. US20030182687A1ton, Kelley C.
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS
FILE REFERENCE: CL1929 US NA
CURRENT APPLICATION NUMBER: US/10/358,917
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: 60/355,939
PRIOR FILING DATE: 2002-02-11
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 497
TYPE: PRT
ORGANISM: Methylobionas sp.16a
3-10-358-917-8

Query Match 100.0%; Score 2598; DB 14; Length 497;
Best Local Similarity 100.0%; Pred. No. 4.6e-251;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSNDNQVIVIGAGLGLSAAISLATAGFSVQLIEKNKDVGGKLNIMTKDGTFFDLGPS 60
|||||
1 MNSNDNQVIVIGAGLGLSAAISLATAGFSVQLIEKNKDVGGKLNIMTKDGTFFDLGPS 60
|||||
61 ILTMHPIEALFTGAGKNMADYVQIQKVEPHRNFFEDGSDVIDICEDARTORRELDKLP 120
|||||
61 ILTMHPIEALFTGAGKNMADYVQIQKVEPHRNFFEDGSDVIDICEDARTORRELDKLP 120
|||||
121 GYVAFORFLDYKNLCTETAGYFAKGLDGFNDLLKPYGLRSLFSDFVRSMDQVRR 180
|||||
121 GYVAFORFLDYKNLCTETAGYFAKGLDGFNDLLKPYGLRSLFSDFVRSMDQVRR 180
|||||
181 FTSDPKLVILNFIKYVSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAV 240
|||||
181 FTSDPKLVILNFIKYVSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAV 240
|||||
241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVVSNMEVIPAMEKLRSPASELK 300
|||||
241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVVSNMEVIPAMEKLRSPASELK 300
|||||
301 KXQRFEPSCGLVHLGLVDRLYPQLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
|||||
301 KXQRFEPSCGLVHLGLVDRLYPQLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
|||||
361 CKTDPAQAPACEIILPHILPDLDPKLLTAEDYSALRERVLVKLRMGLTDLRQHIVT 420
|||||

361 CKTDPAQAPACEIILPHILPDLDPKLLTAEDYSALRERVLVKLRMGLTDLRQHIVT 420
|||||
421 EEWTPDLIDQAKYYSNQSISYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480
|||||
421 EEWTPDLIDQAKYYSNQSISYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480
|||||
481 VTLGQVLVDKIVADLQ 497
|||||
481 VTLGQVLVDKIVADLQ 497
|||||

RESULT 5
US-10-358-917-14
Sequence 14, Application US/10358917
Publication No. US20030182687A1
GENERAL INFORMATION:
APPLICANT: Cheng, Qiong
APPLICANT: No. US20030182687A1ton, Kelley C.
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS
FILE REFERENCE: CL1929 US NA
CURRENT APPLICATION NUMBER: US/10/358,917
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: 60/355,939
PRIOR FILING DATE: 2002-02-11
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Microsoft Office 97
SEQ ID NO 14
LENGTH: 497
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-10-358-917-14

Query Match 49.6%; Score 1289; DB 14; Length 497;
Best Local Similarity 51.9%; Pred. No. 6e-120;
Matches 254; Conservative 85; Mismatches 148; Indels 2; Gaps 2;

7 QRVIVIGAGLGLSAAISLATAGFSVQLIEKNKDVGGKLNIMTKDGTFFDLGPSILTMPH 66
|||||
3 KHIIVIGGLGGSIAIRMAQSGSVLSYEQNNHIGGVKVRHESDGFGLGPSILTMPY 62
|||||
67 IFEALFTGAGKNMADYVQIQKVEPHRNFFEDGSDVIDICEDARTORRELDKLPFTVAQF 126
|||||
63 IFEKLFYSKXQMSDYVTIKRLPHQMSFFPDGTTIDLYEGIKETGQHAILSKQDIEEL 122
|||||
127 QRPFLDYKNLCTETAGYFAKGLDGFNDLLKPYGLRSLFSDFVRSMDQVRRFISDPK 186
|||||
123 QNYLTYTRIDRITKGYFNGLDGLTSLIIRKPHGLNALINDYVHTMQQADIKRISNPY 182
|||||
187 LVEILNFIKYVSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELGVSI 246
|||||
183 LRQMLGYFIKYVSSSYDAPAVLSNKLPHMQQEQGLWYVEGGIHELANALEKLABEGVTI 242
|||||
247 RLDAEVSBEIQODGRACAVKLANGDVLADIVVSNMEVIPAMEKLRSPASELKMMOR-F 305
|||||
243 HPGARVDNIKTYQREVTVGLTGTGFKVADYIIISNMEVIPTKYKLIHLDTORLKLREF 302
|||||
306 EPCSGLVHLGLVDRLYPQLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVAFCPTDP 365
|||||
303 EPASSGYVHLGVACQYQLAHNFFFTENAVLVNQVVEKVLPPDDPTIYLVTNKTGH 362
|||||
366 AQAPAGCEIILPHILPDLDPKLLTAEDYSALRERVLVKLRMGLTDLRQHIVTEYWT 425
|||||
363 TQAPVGYENIKVLPHPYIQ-DQPTTDEDYAFKDKLCKLRMGLTDLRKHIIYEDVMT 421
|||||
426 PLDIOAKYYSNQSISYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPMVTLSG 485
|||||
422 PEDISKNRNRNGALYGVVADKKNKGFPFKESQYFENLIFYVGSVNPGGMPMVTLSG 481
|||||
486 QLVRDKIVA 494
|||||
482 QQVADKINA 490
|||||

RESULT 6

IS-09-815-242-5696
Sequence 5696, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Orlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5696

LENGTH: 439

TYPE: PRT

ORGANISM: Staphylococcus aureus

IS-09-815-242-5696

Query Match 40.0%; Score 1039; DB 9; Length 439;

Best Local Similarity 48.7%; Pred. No. 5.4e-95;

Matches 205; Conservative 80; Mismatches 134; Indels 2; Gaps 2;

2Y 7 QRVIVIGAGLGLSAAISLATAGSPVOLIEKNKVGKLNIMTKDGTFTDLGPSILTMH 66

3 KHIIVIGGLGISAIRMAQSGYSVSLYEQNHIGKVNHRHSDGFGFDLGPSILTMPY 62

67 IPEALFTGAGKNMADYVQIKVEPHEWNEFDSGVIDLCEADFTQRELDKLGPGTYAQF 126

63 IPEKLFYKQMSDYTIKRLPHQWRSFPDGTITDLYEGIKETQHAILSNKDIEL 122

127 QRFLDYKNLCTETAGYFAKGLDGLWLLKPYGFLRSLSLSDVFRSMQGVRRFTSDPK 186

123 QNVLNTRRIDRIETKGVNGLDLSQIIKFKGPNALINDYVHTWQQAIDKRSNPF 182

187 LVEILNFIKVGSSPDAPALMNLPIYQHYGLVYKGMVGMQAGMKAELAVELGVEI 246

183 LRQMLGPIKVGSSPDAPALMNLPIYQHYGLVYKGMVGMQAGMKAELAVELGVEI 242

247 RLDAEVEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSPASSELKQMR-F 305

243 HTGTRVDNIKTYQRVGTGRLDTGEFVKADVIISNMEVITYKYLLHLGTLQRLKLERGF 302

306 BPSCGLVHLGVDRLLPOLAHNFFVSDPREHFDVAFKSHLSDDPTIYVAPCKTDP 365

303 EPASGVYMLGVACQYPOLAHNFFVSDPREHFDVAFKSHLSDDPTIYVAPCKTDP 362

366 AQAPAGCEIILKILPHIDPDKLLTAEDYSALRERVLKLERMGLTDLRQHVITEEYT 425

363 TQAPVGVENIKVLPHIPIQ-DQPTFTEDAKFDKLDKLERMGLTDLRKYIIVEDVNT 421

OY 426 P 426

Db 422 P 422

RESULT 7

US-10-369-493-19519

Sequence 19519, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 19519

LENGTH: 517

TYPE: PRT

ORGANISM: Myxococcus xanthus

US-10-369-493-19519

Query Match 31.1%; Score 809; DB 15; Length 517;

Best Local Similarity 34.9%; Pred. No. 7.5e-72;

Matches 175; Conservative 100; Mismatches 204; Indels 22; Gaps 7;

OY 3 SNDQRVIVIGAGLGLSAAISLATAGSPVOLIEKNKVGKLNIMTKDGTFTDLGPSIL 62

Db 4 STQGERIVVVGAGVGLAAARLAHQGFQVFEKTKQPGRCNRLQVDFWDLGPTIV 63

63 TMHPIFEALFTGAGKNMADYVQIKVEPHEWNEFDSGVIDLCEADFTQRELDKLGPGT 122

64 LMPEVFETFRVNGRIEDVLTLLRCDFNVYVHFRDGSVDTFTSELCAWGRELVERPGS 123

123 YAQQRFLDYKNLCTETAGYFAKGLDGLWLLKPYGFLRSLSLSDVFRSMQGVRRFTSDPK 173

124 YARYLAFLAQGR-----VQYRTSLDHL--VGRNYAGLRDYLSPRVLARIFQVRAHRR 173

174 MDQGVRSFISDPKIVELNFIKVGSSPDAPALMNLPIYQHYGLVYKGMVGMQAGMKA 233

174 MYADVSFFQDERLRRAAMTQTMYLGVSPYASVAVGLLPTTELGVIGWEPKGLVAIPQ 233

234 AMEKLAVELGVEIRLDAEVESEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLR 293

234 ALERLARBEQVRFYGVAPVERILTDGRTGRVLEGEVVEADAVLCNADLPYAEKLLD 293

294 SPASELKQMRFPBSCGLVHLGVDRLLPOLAHNFFVSDPREHFDVAFKSHRLSDDP 353

294 PKATTLKREKELRYTSSGYMLYLGKGRYPPELLHNVVFGDYKGSFDDIFERFRVPEDP 353

354 TIYLVAPCKTDPAQAPAGCEIILKILPHIDPDKLLTAEDYSALRERVLKLERMGLTD 413

354 SFYVNAPTRIDASLAPSGKDALVYLVVPVPHOHDPDLQWKEG-PKVRAPKFPARMABLGP 412

414 LRQHVITEEWTPTLDIOAKYKSNQSGIYGVVADRFKMLG-PKAPQSSSELNLYVFGSSV 472

413 LESDIEVERVFTPDWAGTFTNLARGSAFG-LSONFTQIGPFRPSNQARVKNLFFVGAST 471

473 NPGGMPMTVLSGOLVRDKIV 493

472 QPGLTFLVLSARLVTERLM 492

RESULT 8

US-09-941-947A-32

Sequence 32, Application US/09941947A

Publication No. US20030030528A1
GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia C.
APPLICANT: Cheng, Qiong
APPLICANT: Dicosimo, Deana J.
APPLICANT: Koffas, Mattheos
APPLICANT: Miller, Edward S. Jr.
APPLICANT: Odom, J. Martin
APPLICANT: Picataggio, Steve
APPLICANT: Rouviere, Pierre E.

TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE

FILE REFERENCE: CL1903 US NA

CURRENT APPLICATION NUMBER: US/09/941,947A

CURRENT FILING DATE: 2001-09-01

PRIOR APPLICATION NUMBER: 60/229,907

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,858

PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 60

SOFTWARE: Microsoft Office 97

SEQ ID NO 32

LENGTH: 492

TYPE: PR1

ORGANISM: Pantoea stewartii

S-09-941-947A-32

Query Match 26.4%; Score 687; DB 10; Length 492;

Best Local Similarity 33.3%; Pred. No. 1.2e-59;

Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

y 11 VIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDFTFDLGSILTMHIFEA 70

b 6 VIGAGPGLALAIRLOAGIPVLLLEQRDKPGRAVYVQEQFTFDAGTVITDPSAIEE 65

y 71 LFTGAGKXNADYVQIQKVEPHNFFEDSGSVIDLCEAETQREELDKLPGTVACQFRL 130

b 66 LFLAGKQLXQDYVELLPVTPFYRLCWESKGVNFYNDQALEAQIQFNPDRVAGYRAF 125

y 131 DYSKNLCTETEAGYFAKGLDGFMDLLKFGYPLRS---LLSFDVFRSMQGVRRFISDPKL 187

b 126 DYGRAVNE---GYLKLGTVPF---LSFKDMLRAAPQLAKLQAWRSVYKVGAGYIEDEHL 179

y 188 VEILNTPFYKVGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELGVEIR 247

b 180 RQAFPSHLLVGNPNPATTSIYTLIHALLEREWGVWPPRGGTGALVNGMILKLPQDLGGVW 239

y 248 LDAEVSIEIQDQGRACAVKLANGDVLPAIVVSNMEVIPAMEKILR---SPASELKKMQR 304

b 240 LNAVSHMETVGDKIQAQVLEDGRRFETCAVASNADVHTYRDLLSQHPAAAKQAKKLS 299

y 305 FEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVAECKTD 364

b 300 KRMNSLFLVLYFGLNHHHDQLAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVTD 359

y 365 PAQAPACGEIILKILPHILPDPLKLTAEYSAIRERVLKERMGLTDLRQHIIVTEBYM 424

b 360 PSLAPECGSYVVLAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVTHRMF 419

y 425 TPLDIQAKYYSNGSGSYGVVADRFKNLGFAPORSELSNLYFVGSSVNPGGMPMTLS 484

b 420 TPFDRLNMGQSAFSEVPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGS 479

y 485 GOLVRDKIVADL 496

b 480 AKATAGLMLSDL 491

RESULT 9

US-10-218-118-8

Sequence 8, Application US/10218118

Publication No. US20030148319A1

GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia

APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong

TITLE OF INVENTION: Genes Encoding Carotenoid Compounds

FILE REFERENCE: CL1876 US NA

CURRENT APPLICATION NUMBER: US/10/218,118

CURRENT FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: 60/312,646

PRIOR FILING DATE: 2001-08-15

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Microsoft Office 97

SEQ ID NO 8

LENGTH: 492

TYPE: PR1

ORGANISM: Pantoea stewartii

US-10-218-118-8

Query Match 26.4%; Score 687; DB 14; Length 492;

Best Local Similarity 33.3%; Pred. No. 1.2e-59;

Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

y 11 VIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDFTFDLGSILTMHIFEA 70

b 6 VIGAGPGLALAIRLOAGIPVLLLEQRDKPGRAVYVQEQFTFDAGTVITDPSAIEE 65

y 71 LFTGAGKXNADYVQIQKVEPHNFFEDSGSVIDLCEAETQREELDKLPGTVACQFRL 130

b 66 LFLAGKQLXQDYVELLPVTPFYRLCWESKGVNFYNDQALEAQIQFNPDRVAGYRAF 125

y 131 DYSKNLCTETEAGYFAKGLDGFMDLLKFGYPLRS---LLSFDVFRSMQGVRRFISDPKL 187

b 126 DYGRAVNE---GYLKLGTVPF---LSFKDMLRAAPQLAKLQAWRSVYKVGAGYIEDEHL 179

y 188 VEILNTPFYKVGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELGVEIR 247

b 180 RQAFPSHLLVGNPNPATTSIYTLIHALLEREWGVWPPRGGTGALVNGMILKLPQDLGGVW 239

y 248 LDAEVSIEIQDQGRACAVKLANGDVLPAIVVSNMEVIPAMEKILR---SPASELKKMQR 304

b 240 LNAVSHMETVGDKIQAQVLEDGRRFETCAVASNADVHTYRDLLSQHPAAAKQAKKLS 299

y 305 FEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVAECKTD 364

b 300 KRMNSLFLVLYFGLNHHHDQLAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVTD 359

y 365 PAQAPACGEIILKILPHILPDPLKLTAEYSAIRERVLKERMGLTDLRQHIIVTEBYM 424

b 360 PSLAPECGSYVVLAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVTHRMF 419

y 425 TPLDIQAKYYSNGSGSYGVVADRFKNLGFAPORSELSNLYFVGSSVNPGGMPMTLS 484

b 420 TPFDRLNMGQSAFSEVPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGS 479

y 485 GOLVRDKIVADL 496

b 480 AKATAGLMLSDL 491

RESULT 10

US-10-369-493-234

Sequence 234, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

Sequence 8, Application US/10218118

Publication No. US20030148319A1

GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 234

LENGTH: 491

TYPE: PRT

ORGANISM: Xenorhabdus nematophilus

US-10-369-493-234

Query Match 25.9%; Score 674; DB 15; Length 491;

Best Local Similarity 29.2%; Pred. No. 2.3e-58;

Matches 145; Conservative 111; Mismatches 225; Indels 16; Gaps 3;

8 RVIVIGAGLGSLAAISLATAGFSVOLIENKNDKVGKLNIMTKDGTFTDLGSPILTMPIH 67

3 RIPVIGSGFGIALAIRLQANIPITLLDRKDGGRAYVYQDKGTFDAGPVIITDPSA 62

68 FEALFTGAGKNMADYVQIQKVEPHWRNFEDGSDVIDLCEDAETQRRELKLGPGTVAQFQ 127

63 IEELFLLAGKMSDYIELLPISPFYRLQWQSGEIFDYGGQVVKLEKQIARFNPQVQYR 122

128 RFLDYSKNLCTETAGYFAKGLDGFWDLLKPYGFLRSLLSFDVFRSMDOQVRRFISDPKL 187

123 EFIRYKAV---FTGVLYKQMSDFLSFRSMWCAMPSLVQLSAKSVYGVSKIKNEHL 179

188 VEILNYFYKYVGSSPYDAPALMNLPIYQHYGLWYKGMVYKGMVYKGMVYKGMVYKGMV 247

180 RQAFSPHSLIIGNPFTTSYITLITLREGGVWPRPGTGALVNLVQLFIDIGGVCK 239

248 LDAEVSIEIQKDGACAVKLANGDVLPAIVWSNMEVIPAMEKLL-----RSPASBL 299

240 LNTQVESFITNDQITAKLNDQREIYCKAASNAADVHTYEKLGSHCVGKERSDLKE 299

300 KMQRPFPSCSLVHLGVRLYQPLAHNFFYSDHPREHDAVFKSHRLSDDDPTIYLV 359

300 KCM-----SNSLFVYFGLNKLHQLAHTTICFNRYKELLDIFGNSIPGDFSYLHS 354

360 PKTDPQAQAPAGCEIILKILPHIPLDPPKLLTAEDYSALAEVRLVLEKRMGLTDLRCHIV 419

355 PSVTDPTLAPPCASVYVLPVPHLQNAIWNLEESPLRDLRFSVLEKYIYIPCLRDQVL 414

420 TEYWTPLDIQAKYNSQSGIYGVVADRFRNKGFKAPQRSSELSNLYFVGSGVNPQGM 479

415 TEKIFTPEDFDELNAYHGSFAPIELLQSAWPRPHNRSDIENLYIVGAGTHPGAGIP 474

480 MVLSSGQLVRKIVADL 496

475 GVLASAKITANIMTISDL 491

RESULT 11

US-10-358-917-12

Sequence 12, Application US/10358917

Publication No. US20030182687A1

GENERAL INFORMATION:

APPLICANT: Cheng, Qiong

APPLICANT: No. US20030182687A1, Kelton, Kelley C.

APPLICANT: Tao, Luan

TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS

FILE REFERENCE: CL1929 US NA

CURRENT APPLICATION NUMBER: US/10/358,917

CURRENT FILING DATE: 2003-02-05

PRIOR APPLICATION NUMBER: 60/355,939

PRIOR FILING DATE: 2002-02-11

NUMBER OF SEQ ID NOS: 28

SOFTWARE: Microsoft Office 97

SEQ ID NO 12

LENGTH: 502

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-10-358-917-12

Query Match 25.9%; Score 672; DB 14; Length 502;

Best Local Similarity 31.5%; Pred. No. 3.8e-58;

Matches 162; Conservative 120; Mismatches 191; Indels 42; Gaps 14;

QY 8 RVIVIGAGLGSLAAISLATAGFSVOLIENKNDKVGKLNIMTKDGTFTDLGSPILTMPIH 67

Db 2 KIAVIGAGVTGLAAARIAISQGEHVTIFKKNVVGEMNQLKDGTFDAGPVIITDPSA 61

QY 68 FEALFTGAGKNMADYVQIQKVEPHWRNFEDGSDVIDLCEDAETQRRELKLGPGTVAQFQ 127

Db 62 YKDVFTACGKNVEDYIELRLQYIVDYFQDHDRIITVPTDLAELQOMLESIEPGSTHGF 121

QY 128 RFLD--YSKNLCTETAGYF-----AKGLDGFWDLLKPYGFLRSLLSFDVFRSM---DQ 177

Db 122 SFLTDVYK---VEIARVFLERTYKPSD-----FYN-MTSLVQGAKLKTLNADQL 170

QY 178 VRRFISDFKLVBILNYFYKYVGSSPYDAPALMNLPIYQHYGLWYKGMVYKGMVYKGMVYK 237

Db 171 IERYIDNEKIQLAPQTLYIGIDPKRGPSLYSIIPMIEMMGVHFIKGMVYKGMVYKGMVYK 230

QY 238 LAVELGVEIRLDAEVS--IQKQGRACAVKLANGDVLPAIVWSNMEVIPAMEKLLRSP 295

Db 231 LNKDLGVNIELNABIEGIIIDPKFKRADAIVK--NGDIRKFKILCTADFPSSVAESLMPDF 289

QY 296 ASELK---KMQRPFPSCSLVHLGVLD--RLYPQLAHNFFYSDHPREHDAVFKSHRLS 350

Db 290 APIKKYPPHKLADLDYSCAPLAMYIGIDIDVTQVRLHNVIPSDDFRGNIEEIPFG--RLS 348

QY 351 DDPTIYLVAPCKTDPQAQAPAGCEIILKILPHIPLD-----PKLLTAEDYSALAEVRL 403

Db 349 YDPSIYVYVPAVADKSLAPEGKTIYVIMPTBELKGTSGSDWSDEALTQO-----IKETIY 404

QY 404 VKLERMGL--TDLRCHIVTEYWTPLDIQAKYNSQSGIYGVVADRFRNKGFKAPQRSSEL 462

Db 405 RKLATIEVFEDIKSHVSEITIFENDEQTYHAKFGSAFGLMPLTAQSNYYRPNQWSRDY 464

QY 463 SNLYFVGSGVNPQGMVMTLSGQLVRKIVADILQ 497

Db 465 KDLYFAGASTHPGAGVPIVLTSAKITVDEMIXDIE 499

RESULT 12

US-10-369-493-18644

Sequence 18644, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 18644

LENGTH: 512

TYPE: PRT

ORGANISM: Halobacterium sp. NRC-1

US-10-369-493-18644

Query Match 25.5%; Score 662.5; DB 15; Length 512;

Best Local Similarity 32.4%; Pred. No. 3.5e-57;

Matches 169; Conservative 81; Mismatches 217; Indels 55; Gaps 11;

QY 9 VIVIGAGLGSLAAISLATAGFSVOLIENKNDKVGKLNIMTKDGTFTDLGSPILTMPIH 68

Db 9 VAVIGAGGGSTACYLADAGADYTVVEKTDQICGRASTLERDGRFDMGSPVLMEDVF 68

QY 69 BALFTGAGKNMADYVQIQKVEPHWRNPF-----DGSVIDLCEDAET 110

b 69 EFFAFSDHSDYGLTDLDPHYRFFKONREGRGSDAGLNDVDGDTIDVTDPREQ 128
y 111 QRRELDKLGPGTYAQORFLDYSKNLCTETETAGYFAGLGDG-----WDLKPYG 160
b 129 VQVFDVAVRAGAGDVLDDYLAQAK-----NYEVGMHFVKTRDRPRVDMMDPKLAE 180
y 161 PLRSLSPDVFSSMQGVRRFISDPKLVLLNYFIKYVSSPYDAPALMNLPLPIQHYG 220
c 181 YARGL---TLGSMQCHVEQYFDHFKLQVQMGYTLVFLGSGPDTTPALYNLMSHVDPLG 237
y 221 LMYVKGMYGMAQAMEKLAVELGVIRLDAEVEITQKQGRACAVKLAN--GDVLPADIV 278
c 238 VYVPEGMGVVDGTALARELGVEFGTCHPUSAIK---GRGGFKLDTPEGDAVLADV 294
y 279 VNMVEVIPAMEKL-----RSPASELKNQORPEPSCGLVHLGVDRLYPOLAHNFPYSD 334
b 295 VSDADYAHTEQALLPAQRQYDADYDAP--SAFLLYLGVGEVPELAHHTLVLP 352
y 335 HPREHFDVAFKSHRLSDDDPTIYLVAPCKTDPAAQAG--CEIIKILPHIPHPDKLLTAE 393
c 353 DWDTHFAQIFDDPAMPEDDPAYVLCVPSKTDITVAPDGHSLPALVPVAPGLD-DTPAVRE 411
y 394 DYSALRERVVLKRMGLDRLQHIIVTEYTPDLIOAKYVSNQSGIYGVVADRFKNLGF 453
c 412 SY---RDLVLDIAENTGVLDRLDRIWVEERFSVSEFADRYNSHQGTALGLAHTLRQALF 468
y 454 KAPQSSSELNLYFYVGGSVNPGGGMWVTLSCQVLRDKIVAD 495
c 469 RPPHRSALDGLYFTGANTPDTGIVPMCLISGEVTANVLAD 510

RESULT 13
S-10-369-493-18983
Sequence 18983, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT FILING DATE: 2003-02-28
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 18983
LENGTH: 499
TYPE: PRT
ORGANISM: Anabaena FCC7120
S-10-369-493-18983

Query Match 25.2%; Score 654; DB 15; Length 499;
Best Local Similarity 32.2%; Pred. No. 2,4e-56;
Matches 158; Conservative 112; Mismatches 187; Indels 34; Gaps 12;
y 6 NORVIVGAGLGLSAAISLATAGFSVQIIEKNKVGKLMIMTKDGTDFDLGSPILTMP 65
c 2 SKKVALVGAGPGGLATRLAGLGQVBEFAEAEVGVGRMGPEVDSVAPDTGFTILQLP 61
y 66 HIFELFTGAGKNMADYVQIQKVEHNRNFFEDSGVIDLCEADATQRELDKLGPGTYAQ 125
c 62 HUYLEFEEAGLNFAFYQLKLEPYTLKFWMDGTQDLTDSQFTQATLSDPLA 121
y 126 FORFLDYSKNLCTETETAGVF-----AKGLDGFW---DLLKPYGLRSLLSFQVFRSMQ 176
c 122 FDRW--YSEHI-RKVELGKPYLAGPARSIFGLYLPDELKMF-----LSFRFWENLYQ 171
y 177 GYRFRISDPKLVLLNYFIKYVSSPYDAPALMNLPLPIQHYGLWYVKGMYGMAQAME 236

Db 172 HWRRFQDRLVVALYSKYLGMPTVASSVPSLPTLEFSQGVHHPVGGFRALAOGLA 231
Qy 237 KLAVELGVIRLDAEVEITQKQGRACAVKLANGDVLDPADIVSNMEVIPAMEKLLRSPA 296
Db 232 NAAQDLGVXILHLSPVHQIWIQGGVGRGLELADASHRQFDTTVINADFAVAVRHLLPTSA 291
Qy 297 SEL---KNQORPEPSCGLVHLGVDRLYPOLAHNFFYSDPREHFDVAFKSHRLSD-D 352
Db 292 RGRYTDNKLGMQFSCSPMLYLGINRYEDLPHQIYLSNIRLERPWWDDSDALDET 351
Qy 353 PTIYLVAPCKTDPAAQAGCEIILPHIPH---LDPEKLLTAEVYSALRERVIVKLER 408
Db 352 PPFVVCNPTIIDESNAPAGHSTLFLVLPINTSVAYDWD--IKQKSYT---DFILKRLHL 406
Qy 409 MGLTDLRQHIVTEYTPDLIOAKYVSNQSGIYGVVADRFKNLG-PKAPQSSSELNLYF 467
Db 407 LGVHNIEQHIVTQSCYTAQSWLDDRYVHLGAVFN-LSHNLTQLGPPRPPIRSENTAGLYW 465
Qy 468 VGGSVNPGGGM 478
Db 466 IGGAVHFGSGL 476

RESULT 14
US-10-369-493-20438
Sequence 20438, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT FILING DATE: 2003-02-28
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20438
LENGTH: 498
TYPE: PRT
ORGANISM: Rhodospseudomonas palustris
US-10-369-493-20438

Query Match 24.8%; Score 645; DB 15; Length 498;
Best Local Similarity 30.8%; Pred. No. 1,9e-55;
Matches 153; Conservative 103; Mismatches 218; Indels 22; Gaps 8;
Qy 10 IVIGAGLGLSAAISLATAGFSVQIIEKNKVGKLMIMTKDGTDFDLGSPILTMPHIFE 69
Db 14 VVIGSGFGGLAAVRLGAKYRVTVLEKLDKAGGRAYVHKQDGSFSDAGPTIVTAPYIFE 73
Qy 70 ALFTGAGKNMADYVQIQKVEHNRNFFEDSGVIDLCEADATQRELDKLGPGTYAQFOR 129
Db 74 ELWKLCGKMSDDITLKPMSPPIRFDGCHFDVSDRDAVLDOIAKFCDDVPAYDRF 133
Qy 130 LDYSKNLCTETETAGYFAGK---LDGFWDLKPYGLRSLLSFQVFRSMQGVRRFRISPK 186
Db 134 MAASHEI---PKVQPEQLCDQPSHFTDMLKI---APAMIKLESYRSYVGLVAKHFKDPK 187
Qy 187 IVEILNYFIKYVSSPYDAPALMNLPLPIQHYGLWYVKGMYGMAQAMEKLAVELGVEI 246
Db 188 LRQVFSFHLGLTGPNFMSVYCIITYLEKQMGVHSAMGTTGALVTGLVNLISQCGNTI 247
Qy 247 RLDAEVEITQKQGRACAVKLANGDVLDPADIVSNMEVIPAMEKLLRSPASELK-----K 301
Db 248 RYNQDVRQIVVENGTACGKVLADGEVIKADIVSNADSASTRVLL---PPETRKRWTDAX 305
Qy 302 MGRPEPSCGLVHLGVDRLYPOLAHNFFYSDPREHFDVAFKSHRLSDPTIYLVAPC 361

Db 306 TEKSYNSLFWVPGTKRRYEDVKHHTILGPRYKELISDIFSRKVVAEDFSLYLRPT 365
Qy 362 KTDPAQAGCELIKILPHIHL--DPKLLTADYSALRSERVLKRLMGLTDLRQHV 419
Db 366 ATDPSLAPQGGDTTFVLSVFNLLGDDTWHYKAETI---RASAXMLGATVLPDLNQIA 422
Qy 420 TEYWTPLDIOAKYNSOGSYGVVADRPKNLGFAPQSRSELSNLYFVGGSVNPQGGMP 479
Db 423 TSKITPTIDFQRLSSFGAAGLEPVLWQSAWFRPHNQSEDVKNLYLVGAGTHPGAGLP 482
Qy 480 MVTLSGQLVRDKIVADL 495
Db 483 GV-ISSARVLDALVPE 497

RESULT 15
US-09-547-267-5
Sequence 5, Application US/09547267
Patent No. US20020147371A1
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Passamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-547-267-5

Query Match 24.5%; Score 637; DB 9; Length 494;
Best Local Similarity 31.2%; Pred. No. 1.2e-54;
Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;
Qy 10 IVIGAGLGGSLAATAGFSGVOLIEKNDKVGKLNIMTKDGFDFDLGFSILTMPIHFE 69
Db 5 IVIGAGFGGLALATLQSGAGIATTVARADKGGRAYVWMDQGHVFDAGTVVTDPSLR 64
Qy 70 ALFTGAGKNMADYQIQKVEPHRNFFPDGSDVLDCEDAETQRRELKLGPGTYAQRPF 129
Db 65 ELWALSGQPMERDVTLLPVSFFRLTWADGSRFEYVANDDDDELIRQVASFNPADVDGYRRF 124

Qy 130 LDYSKNLCETEAGYFAKGLDGFWDLLKFGYCPLESLLSFDVFRSMDQGVRRFISDPKIVE 189
Db 125 HDYAEVYRE--GYLKLGTTPFLKLGMLNAAAPALMRLQAYRSVHSMVARFIQDPHLRQ 181
Qy 190 ILNYFIKYVSSPYDAPALMMLLPYIOYHYGLWYVKGMYGMAQAMEKLAVELGVRLD 249
Db 182 APSFHTLLVGNPNPFTSSIVALLHALERRGGVWFAKGGTQNLVAGMVALFERLGGTLLN 241
Qy 250 AEVSEIQQDGRACAVKLANGDVLPAIVVSNMEVI PAMEKLL-----RSPASELKQM 303
Db 242 ARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAAILNR-Q 300
Qy 304 RPEPSCGLVHLGVDRLYPOLAHNFFYSOHPREHFDAYPKSHRLSDDPITIVLVAPECT 363
Db 301 RW--SMSLFLVHLFGLSKRPENLAHHSVIFOPRYKGLVNBIFNGRPLDDDFSMYLSHSPCVT 358
Qy 364 DPAQAPAGCEIIKILPHIHL---DPKLLTADYSALRSERVLKRLMGLTDLRQHIVT 420
Db 359 DPSLAPEGMSTHYVLAAPVPHLGRADVDEAEAPGYA--ERIFEELERRAIPDLRKHITV 415
Qy 421 EBYWTPLDIOAKYNSOGSYGVVADRPKNLGFAPQSRSELSNLYFVGGSVNPQGGMP 480
Db 416 SRIFSPADFSSTLSAHGSAFVSVEPILTSQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
Qy 481 VTLGQLVRDKIVADL 496
Db 476 VVGSAKATAQVMLSOL 491

Search completed: February 29, 2004, 15:27:52
Job time : 34.998 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:35:44 ; Search time 15.1029 Seconds
(without alignments)

1698.885 Million cell updates/sec

File: US-09-941-947A-24

Effect score: 2598
Sequence: 1 MNSNDNRQVIVIGAGLGGLS.....MEMVTLSGQLVRDKIVADLQ 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgm2_6/ptodata/2/iaa/5A.COMB.pdp:*
2: /cgm2_6/ptodata/2/iaa/5B.COMB.pdp:*
3: /cgm2_6/ptodata/2/iaa/6A.COMB.pdp:*
4: /cgm2_6/ptodata/2/iaa/6B.COMB.pdp:*
5: /cgm2_6/ptodata/2/iaa/ECTUS.COMB.pdp:*
6: /cgm2_6/ptodata/2/iaa/backfiles1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2598	100.0	497	4	US-09-934-903-18
2	703.5	27.1	489	1	US-08-095-726-10
3	703.5	27.1	489	1	US-08-096-043-10
4	703.5	27.1	489	1	US-08-096-623A-10
5	702.5	27.0	489	1	US-08-095-726-8
6	702.5	27.0	489	1	US-08-096-043-8
7	702.5	27.0	489	1	US-08-096-623A-8
8	671	25.8	492	1	US-07-783-705A-4
9	637	24.5	494	3	US-08-660-645A-5
10	637	24.5	494	3	US-09-298-718-5
11	637	24.5	494	3	US-09-546-969-5
12	637	24.5	494	3	US-08-980-832-4
13	637	24.5	494	4	US-09-547-267-5
14	637	24.5	494	4	US-09-920-923B-4
15	618.5	23.8	582	4	US-09-091-725-17
16	607.5	23.4	511	4	US-09-934-903-16
17	596	22.9	290	4	US-08-936-165A-462
18	184.5	7.1	610	4	US-09-443-184-55
19	155.5	6.0	508	2	US-08-472-028A-4
20	155.5	6.0	508	2	US-08-808-931-4
21	155.5	6.0	508	3	US-08-808-333-4
22	155.5	6.0	508	3	US-09-050-603A-4
23	155.5	6.0	508	3	US-09-102-420B-4
24	155.5	6.0	508	3	US-09-071-296-4
25	155.5	6.0	508	3	US-09-196-268-4
26	155.5	6.0	508	3	US-09-015-683-4
27	155.5	6.0	508	4	US-09-191-998-4

28	155.5	6.0	508	4	US-09-497-698-4	Sequence 4, Appli
29	150.5	5.8	598	3	US-09-352-159-42	Sequence 42, Appl
30	149.5	5.8	578	4	US-09-690-942-10	Sequence 10, Appl
31	144	5.5	524	3	US-09-230-388-1	Sequence 1, Appli
32	144	5.5	524	4	US-09-912-176-1	Sequence 1, Appli
33	143.5	5.5	571	4	US-09-690-942-15	Sequence 15, Appl
34	143	5.5	598	3	US-09-352-159-40	Sequence 40, Appl
35	143	5.5	991	3	US-09-352-159-27	Sequence 27, Appl
36	143	5.5	991	3	US-09-352-168-27	Sequence 27, Appl
37	143	5.5	1000	3	US-09-352-159-25	Sequence 25, Appl
38	143	5.5	1000	3	US-09-352-168-25	Sequence 25, Appl
39	143	5.5	1196	3	US-09-352-159-31	Sequence 31, Appl
40	143	5.5	1196	3	US-09-352-168-31	Sequence 31, Appl
41	143	5.5	1205	3	US-09-352-159-29	Sequence 29, Appl
42	143	5.5	1205	3	US-09-352-168-29	Sequence 29, Appl
43	141.5	5.4	462	3	US-09-352-159-8	Sequence 8, Appli
44	141.5	5.4	462	3	US-09-352-168-8	Sequence 8, Appli
45	141.5	5.4	829	3	US-09-352-159-33	Sequence 33, Appl

RESULT 1
US-09-934-903-18
; Sequence 18, Application US/09934903
; Patent No. 6660507
; GENERAL INFORMATION:
; APPLICANT: Koffas, Mattheos
; APPLICANT: Odom, J. Martin
; APPLICANT: Schenzle, Andreas J.
; APPLICANT: No. 6660507ton, Kelley C.
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Rovierte, Pierre
; APPLICANT: Picataggio, Stephen
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
; FILE REFERENCE: CL1646 US NA
; CURRENT APPLICATION NUMBER: US/09/934, 903
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229, 907
; PRIOR FILING DATE: September 1, 2001
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ORF9
US-09-934-903-18
Query Match 100.0%; Score 2598; DB 4; Length 497;
Best Local Similarity 100.0%; Pred. No. 8.8e-259;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNSNDNRQVIVIGAGLGGLSAAISLATAGSFVQLIEKNDKVGKGLNIMTKDGFDFDLGSP	60
Db	1	MNSNDNRQVIVIGAGLGGLSAAISLATAGSFVQLIEKNDKVGKGLNIMTKDGFDFDLGSP	60
Qy	61	ILTMTHIPEALFTGAGKNADYVQIQKVEPHWRAPFFDGSVIDLCEDAETORRELDKGP	120
Db	61	ILTMTHIPEALFTGAGKNADYVQIQKVEPHWRAPFFDGSVIDLCEDAETORRELDKGP	120
Qy	121	GTIAQFORFLDYSKNLCTETETAGTFAGKLDGFWDLKFPGLRSLSPDFVFRSMDDQVRR	180
Db	121	GTIAQFORFLDYSKNLCTETETAGTFAGKLDGFWDLKFPGLRSLSPDFVFRSMDDQVRR	180
Qy	181	FISDPKLEIILNFKYVSSPYDAPALMNLIPYQHYGLWYVKGMYGMAQAMEKLAV	240
Db	181	FISDPKLEIILNFKYVSSPYDAPALMNLIPYQHYGLWYVKGMYGMAQAMEKLAV	240
Qy	241	ELGVEIRLDAEYSEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLRSASELK	300

ALIGNMENTS

db 241 ELGVIRKDAEVESEIQKQGRACAVKANGDVLPAIVVSNMEVIPAMEKLRSPASELK 300
iy 301 KMORFEPSCGLVHLGVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
b 301 KMORFEPSCGLVHLGVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
y 361 CKTDPAQAPAGCEIILPHIPHLDPKLLTADYSALRERVVLKLERMGLTDLRQHI 420
b 361 CKTDPAQAPAGCEIILPHIPHLDPKLLTADYSALRERVVLKLERMGLTDLRQHI 420
y 421 EBYWTPLDIOAKYYSNQGSIYGVVADRFKNLGPAPORSSLSNLYFVGGSNPGGM 480
b 421 EBYWTPLDIOAKYYSNQGSIYGVVADRFKNLGPAPORSSLSNLYFVGGSNPGGM 480
y 481 VTLSGQLVRDKIVADLQ 497
b 481 VTLSGQLVRDKIVADLQ 497

RESULT 2

S-08-095-726-10

Sequence 10, Application US/08095726

Patent No. 5530189

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B

TITLE OF INVENTION: Beta-Carotene Biosynthesis in
Genetically Engineered Hosts

NUMBER OF SEQUENCES: 79

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/095,726

FILING DATE: 21-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/785,566

FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, No. 5530189val B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 3128567180

TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 489 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-095-726-10

Query Match 27.1%, Score 703.5; DB 1; Length 489;
Best Local Similarity 32.9%; Pred. No. 1.5e-63;
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

2y 7 ORVTIVIGAGLGAASLATAGSVOLIEKNDKVGKLMIMTKQGTFFDLGPSILTM 66

db 2 EKTVVIGAGLGAALATLQAGIPTVLLQRDKPGGRAYVWHDQGTFFDAGFTVIT 61

Qy 67 IFEALFTGACKOMADYVOIQKVEPHRNFEEDSGSVLDLCEDAETQRRELDKLGPGTYAOF 126
Db 62 ALFALFTLAGRMEDVRLLPVAPFYLRLCWESGKTLDYANDSFLEBAQITQPNPROVEGI 121
Qy 127 QRFLDYSKNLCTEAGYFAKG---LDGFWDLKFKYGLRSLSPDFVFRSMDQGVRRFIS 183
Db 122 RRFAYSQAVFQF---GYLRIGSVPLSLFRDMLR-AGP---QLLKQAWQSVQSVSRFIE 175
Qy 184 DPKLVEILANFYKYVSSPYDAPALMNLPLPYQYHGLVYVKGWYGMQAQAMEKLAVELG 243
Db 176 DEHLRQAFPSHLLVGNFTTSSIYTLHAUSERWGVFPFGGTGALVNGWVLFLLFDLG 235
Qy 244 VEIRLDAEVESEIQKQGRACAVKANGDVLPAIVVSNMEVIPAMEKLRSPASELK-- 300
Db 236 GEIELMARVEELVADNRVSQVELADGRIFDTPDAVASNADVNTYKLLGTIPVQKRAA 295
Qy 301 KMORFEPSCGLVHLGVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360
Db 296 RLBRKMSNSLFLVLYFGLNQPHSLAHTTICFGPRYRELIDEIFTGSAADDFSLYLHSP 355
Qy 361 CKTDPAQAPAGCEIILPHIPHLDPKLLTADYSALRERVVLKLERMGLTDLRQHI 420
Db 356 CVTDFSLAPPPCASFYVLPVPHLGNAPLDWQEGPKLRDRIFDYLEERYMPLGRSLQVLT 415
Qy 421 EBYWTPLDIOAKYYSNQGSIY-----GVVADRFKNLGPAPORSSLSNLYFVGGSV 472
Db 416 QRIFTRQTSRHWAILGLSLFIEPPSLTQGLFA-----ANATRHSNLYLVAAGT 464
Qy 473 NPGGMPMTLSGQLVRDKIVADLQ 497
Db 465 HPGAGIPGVVGLAEASTASLMIEDLQ 489

RESULT 3

US-08-096-043-10

Sequence 10, Application US/08096043

Patent No. 5530189

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B

TITLE OF INVENTION: Lycopen Biosynthesis in
Genetically Engineered Hosts

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/096,043

FILING DATE: 22-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/785,568

FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, No. 5530189val B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 3128567180

TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 10:

465 HPGAGIPGVVGLAESTASLMIEDLQ 489

RESULT 5

US-08-095-726-8
Sequence 8, Application US/08095726
Patent No. 5530188

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Beta-Carotene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530188val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-095-726-8

Query Match 27.0%; Score 702.5; DB 1; Length 489;
Best Local Similarity 32.9%; Pred. No. 1.9e-63;
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

7 QRVIVIGAGLGSAALSLATAGPSVQLIEKNDKVGKLNIMTKDFTDLGPSILTMHP 66
2 KKTIVIGAGFGLALALRLQAGIPTVLLQQRDKPGRAVWHDQGTFDAGFTVIDPT 61
67 IFEALFTGAGNMADYVQIQKVEPHNRNFFEDSGVIDLCEDAETQRRLDKLPGTYAQF 126
62 ALEALFTLAGRMEDYVRLLPVKPFYRLCWESGKTLDYANDSFELEAQITQPNPRDVEGY 121
127 ORFLDYSKNLCTEAGYFAGK--LDGFWDLLKFGYGLRSLSLSPDVFPSMDQGVRRFIS 183
122 RRLFLAYQAVPQE--GYRLGVSVPFTLSFDMLR-AGP--QLLKLAQWSVQVSURFIE 175
184 DPKLVEILNYFIKTVGSSPDAPALMNLPIYIQHVLGWLTVKGYMGMAQAEKLAVELG 243
176 DEHLRQAFPSHLLVGNPFTTSIYTLIALEREWGVWPEEGGTGALVNGMYKLFDTLG 235
244 VEIRLDAVSEIQKQDGRACAVKLANGDVLPAITVSNMMEVIPAMEKILLES-PASELK-- 300
236 GEIELNARVESLVDADRNVSRQLADGRIFDTDAVASNADVNNTYKLLGTIPVQGRAA 295

301 KMORFSPSCSLVHLGVDRLYPOLAHNFFYSDHREHFDVAFKSHRLSDDDPTIYLVA 360
296 RLEKKSMSNSLFLVYGLNQPHQSLAHHTICGPRYRELIDEIFTSALADDFSLYHSP 355
361 CKTDPAAQAPACBIIKILPHIPHLDPKLLTADYSALRERVVLKLERMGLTDLROHIV 420
356 CVTDPSLAPPCCASFVYLAPVPHLGNAPLDWAQEGPKLRDRIFDYLEERYMPLGRSOLVT 415
421 EYMTPLDIQAKYVNSGSIY-----GVVADRFKNLGFKAPQPSSELSNLYFYGGSV 472
416 QRIFTRQTSRHAWIILGSLFIEPPSLTQOLFA-----ANATHSNLYLVAAGT 464
473 NPGGEMVMTLSGQLVRDKIVADLQ 497
465 HPGAGIPGVVGLAESTASLMIEDLQ 489

RESULT 6

US-08-096-043-8
Sequence 8, Application US/08096043
Patent No. 5530189

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Lycopene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,043
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,568
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-096-043-8

Query Match 27.0%; Score 702.5; DB 1; Length 489;
Best Local Similarity 32.9%; Pred. No. 1.9e-63;
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

7 QRVIVIGAGLGSAALSLATAGPSVQLIEKNDKVGKLNIMTKDFTDLGPSILTMHP 66
2 KKTIVIGAGFGLALALRLQAGIPTVLLQQRDKPGRAVWHDQGTFDAGFTVIDPT 61
67 IFEALFTGAGNMADYVQIQKVEPHNRNFFEDSGVIDLCEDAETQRRLDKLPGTYAQF 126

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:33:49 ; Search time 43.7446 Seconds
(without alignments)
3037.172 Million cell updates/sec

tie: US-09-941-947A-38

Effect score: 2768

Sequence: 1 MSAPDAVVGSGHNLVSA.....ALSDSRGKASQMMRRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

SPTREMBL 25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	859.5	31.1	511	16	Q9RY57
2	688	24.9	542	16	Q55808
3	660.5	23.9	565	16	Q8YQ55
4	624.5	22.6	581	4	Q8YRQ1
5	624.5	22.6	581	4	Q8N2H3
6	572.5	20.7	556	10	Q8YVK3
7	562.5	20.3	647	10	Q9FGZ1
8	557.5	20.1	521	16	Q98FP6
9	538.5	19.5	543	16	Q9A3T1
10	537.5	19.4	519	16	Q98FE7
11	533	19.3	518	17	Q97W24
12	525	19.0	517	17	Q974W1
13	509	18.4	539	16	Q8PJ01
14	497	18.0	545	16	Q89FW3
15	469	16.9	524	16	Q98BS8
16	466.5	16.9	544	5	O61196

17	443	16.0	544	16	Q82PC2	Q82pc2 streptomyc
18	441.5	16.0	539	16	Q9L187	Q9l187 streptomyc
19	439.5	15.9	538	17	Q9YCC0	Q9ycc0 aeropyrum p
20	418	15.1	489	16	Q9RYF5	Q9ryf5 deinococcus
21	385.5	13.9	540	16	Q9K414	Q9k414 streptomyc
22	363.5	13.1	523	16	Q98BS6	Q98bs6 rhizobium 1
23	359	13.0	472	2	Q9EY15	Q9ey15 streptomyc
24	349.5	12.6	482	16	Q8FGA2	Q8fsa2 corynebacte
25	348	12.6	489	2	Q9KHE1	Q9khe1 streptomyc
26	340	12.3	472	16	Q9Z8G4	Q9zb84 streptomyc
27	336.5	12.2	535	16	Q8XYC4	Q8xyc4 ralstonia s
28	335	12.1	472	16	Q82LS4	Q82ls4 streptomyc
29	318.5	11.5	536	16	Q7TVK4	Q7tvk4 mycobacteri
30	314.5	11.4	536	16	Q07794	Q07794 mycobacteri
31	309	11.2	533	16	Q82NN7	Q82nn7 streptomyc
32	302	10.9	499	2	Q9K566	Q9k566 mycobacteri
33	296.5	10.7	471	16	Q8NT31	Q8nt31 corynebacte
34	293.5	10.6	471	16	Q82NF9	Q82nf9 streptomyc
35	284	10.3	473	16	Q06826	Q06826 mycobacteri
36	284	10.3	473	16	Q7U032	Q7u032 mycobacteri
37	283	10.2	480	16	Q53244	Q53244 mycobacteri
38	276.5	10.0	588	2	Q83X01	Q83x01 streptomyc
39	275	9.9	523	16	Q9RIX9	Q9rix9 streptomyc
40	274	9.9	507	2	Q9EXL0	Q9exl0 streptomyc
41	273	9.9	480	16	Q7TXH1	Q7txh1 mycobacteri
42	270	9.8	501	16	Q55455	Q55455 synchocyst
43	264.5	9.6	517	2	Q9ZGE4	Q9zge4 heliobacill
44	262.5	9.5	520	16	Q7V6U2	Q7v6u2 prochloroco
45	261	9.4	513	16	Q93HP2	Q93hp2 streptomyc

ALIGNMENTS

RESULT 1

Q9RY57 PRELIMINARY; PRT; 511 AA.
ID Q9RY57
AC Q9RY57
DT 01-MAY-2000 (TRENBL 13, Created)
DT 01-MAY-2000 (TRENBL 13, Last sequence update)
DT 01-JUN-2003 (TRENBL 24, Last annotation update)
DE Phytoene dehydrogenase, putative.
GN DR0093.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1", 1577 (1999).
RL Science 286:1571-1577 (1999).
DR EMBL; AE001872; AAF09686.1; -.
DR PIR; E75561; E75561.
DR TIGR; DR0093; -.
DR InterPro; IPR008151; Phyt_n dehydro.
DR ProDom; PD139017; Phyt_n dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 511 AA; 56527 MW; 7DCC3F81D79BEB9CD CRC64;

Query Match 31.1%; Score 859.5; DB 16; Length 511;
Best Local Similarity 39.7%; Pred. No. 1.9e-45;
Matches 207; Conservative 80; Mismatches 208; Indels 27; Gaps 11;

Y 6 DAVVVGSGHNAALYARABEGHNSVEVLEKDTVLGGAVSTVERFCYKVKVDRGSSAHLIR 65
 S DLIVMGAGHNAALYARABEGHNSVEVLEKDTVLGGAVSTVERFCYKVKVDRGSSAHLIR 64
 Y 66 HSGIIEELGAGHGLRYIDCDPWAFAPAPGTDGPGIVF-HRDLDATCQSI-ERACGKND 123
 b 65 MTFIVRELETRUGLHYLEVDPMFHA-----SDGETPFIHRDAGRTIRELDEKPPG--Q 117
 Y 124 ADAYRRPVAWVSRSRHHVMAKAFSTPTGSLNIGAFGLATARG-----NSLSRQFLAPG 178
 b 118 GDAYGRFLDWTFFARAVADLNFSAFGPLDL-----GKVMRSQQKDMNEQLPR-ILRPY 172
 Y 179 DALLDEYDFSEALKAAALAWFGAGSGPMSEPGTAPMVGFPAALMHVLPDGRVAGSGALSA 238
 b 173 GDVAREYFSERVAPLITWMAAGSGPPDPLSAPFLLWHPLHLYHEGTVARPKGSGGLTK 232
 Y 239 ALASRAVADGATVALGDGVTISR-RNSNHWTTTSGREHVKHVIAGCHILTTLDLGN 297
 b 233 ALARATEAGEGVEFTDAPVKELVKDGAQIGLESGETYTTARAVVGVHILTTANAL-- 290
 Y 298 GGFDRITLTHWRKRIRVGPICGAVLELATSALPSVRGDATTRESTSGQLLVSDRAHLRT 357
 b 291 ---PAEYVPSAARNVRVGVGFGMLRLALSEKVKYR-HHTEPDSRIGLGLIKNERQIMQ 346
 Y 358 AHGAALAGELPPRAPVLGMSFGCDPTIAPAGRHQVTLMSQVRLSLGHRDVAABAE 417
 b 347 GYGEYLAGOFTDPLPLVAMSFSVAVDSDSLAPPNGDVLMLWAQYYPFELA-TGSWETRTAEA 405
 Y 418 ADRIVGEMAFAGFTSDVLDRPIOTPRDIESLGMIGNVHMSLDQMLWRLPEL 477
 b 406 RENILRAFHYAETGRTIVGELVQTPQWLETNLGHRGVHLEMSFDQMPSPFENLKA 465
 Y 478 SGHRVPGADGLYLTASTHPGGVSGVSGASRSAAIALSDSR 519
 b 466 SQVRWPGVQGLYLTASTHPGGGIMGASGRNAARVIVKDLTR 507

RESULT 2

55808 PRELIMINARY; PRT; 542 AA.
 C Q55808; MEDLINE=96127529; PubMed=8590279;
 T 01-NOV-1996 (T-EMBLrel. 01, Created)
 T 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 E Phytocene dehydrogenase.
 N CRTE OR SLR0088.
 S Synechocystis sp. (strain PCC 6803).
 C Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 X NCBI_TaxID=1148;
 N [1]
 P SEQUENCE FROM N.A.
 X MEDLINE=96127529; PubMed=8590279;
 X Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 A Sugiyama M., Tabata S.,
 T "Sequence analysis of the genome of the unicellular cyanobacterium
 T Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 T region from map positions 64% to 92% of the genome.";
 L DNA Res. 2:153-166(1995).
 N [2]
 P SEQUENCE FROM N.A.
 X MEDLINE=97061201; PubMed=8905231;
 A Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 A Miyajima N., Hirose M., Sugiyama M., Sasamoto S., Kimura T.,
 A Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 A Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 A Tabata S.,
 T "Sequence analysis of the genome of the unicellular cyanobacterium
 T Synechocystis sp. strain PCC6803. II. Sequence determination of the
 T entire genome and assignment of potential protein-coding regions.";
 L DNA Res. 3:109-136(1996).
 R EMBL: D64004; BAA10561.1; --
 R F01; S76617; S76617
 R PIR; S76617; S76617
 R GO: GO:0006118; P:electron transport; IEA.

DR InterPro; IPR000759; Adirndx_reductase.
 DR InterPro; IPR000205; NAD_BS.
 DR InterPro; IPR008151; Phyt_n dehydro.
 DR PRINTS; PR00419; ADXRDITASE.
 DR ProDom; PD139017; Phyt_n dehydro; 1.
 KW Complete proteome.
 SQ SEQUENCE 542 AA; 59400 MW; CD673A75E5CE7928 CRC64;
 Query Match 24.9%; Score 688; DB 16; Length 542;
 Best Local Similarity 31.9%; Pred. No. 9.7e-35;
 Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;
 QY 6 DAVVVGSGHNAALYARABEGHNSVEVLEKDTVLGGAVST-----VERFCYKVKVDRGSSAH 61
 b 5 DVVLIAGHNGLYCAAYLLQRLGVLLEKREVPFGAATTEALMPELSQPFENRCAIDH 64
 QY 62 LMRHSGITBELGAGHGLRYIDCDPWAFAPAPGTDGPGIVFHRDLDATCQSIERACT 121
 b 65 EFIFLPGVLQELNLAQYGLYLEYFCDFSVF---CPGLDGAQFMSYRSLEKTAHI-ATYSP 120
 QY 122 KDADAYRRPVAWVSRSRHHVMAKAFSTPTG-----SNLIGAFGLATARGNSLS 171
 b 121 RDAEKTRQFVNYWTDLLNAVQPAFAPFPQALLDLALNYGWNLSVLAGSKTKALDFI 180
 QY 172 RQFLAPGDALLDEYDFSEALKAAALAWFGAGSGPMSEPGTAPMVGFPAALMHVLPDGRVAG 231
 b 181 RTMIGSPEDVLNEWDFSRVKAPLARLCSIEIGAPFSQKSGSSGMMVMAMRHLEGIARPKG 240
 QY 232 GSCALSAALASRAVADGATVALGDGVTISR-RNSNHWTTTSGREHVKHVIAGCHILT 290
 b 241 GTGALTEALVKLQAQGGKILTDITQVRVLVENNQAI GVEVANGSQYRAXK-----GVIS 295
 QY 291 TLD-----LLNGGGFDRITLTHWR-RIRVGPICGAVLRL--ATSALPSYRGDATTRE 340
 b 296 NIDARLFLQLVEPGALAKVQNQNLGERLERRTVNNNEALIKIDCALSGLPHTAMAGP-E 354
 QY 341 STSGLOLVSDRAHLRTAHGAALAGELPPRAPVLGNSP-SGIDPTIAPAGRHQVTLASQ- 398
 b 355 DLTGTTLIADSVRHEEHALIAGQIPDANPSLVIDIPTVLDPNMAPPGQR--TLWIEF 412
 QY 399 WQPYRLSG-----HRDWA-SVAEADRIYGEAEAFAPGFTSDVLDRFIOTPRDISE 450
 b 413 FAFYRIAGLEGTGLMGTTGWTDELKENVADRIDKLTVDAPNLKSLIIGRRVBSPAELAQ 472
 QY 451 LGMIGNVHMSLDQMLWRLPELSCHRVPGADGLYLTGASTHPGGVSGASGRSAA 510
 b 473 LGSYNGVYHLDMSLDQMLWRLPELPIANIYQTP-ITKNLYLTGAGTHPGGSGMPCRNCA 531
 QY 511 RIALSDSR 519
 b 532 RVFLKQRR 540
 RESULT 3
 Q8YQ55 PRELIMINARY; PRT; 565 AA.
 AC Q8YQ55; MEDLINE=21595285; PubMed=11759840;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical protein A113744.
 GN A113744.
 GN Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,


```

11 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
12 Hypothetical protein FLJ90784.
13 Homo sapiens (Human).
14 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
15 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
16 NCBI_TaxID=9606;
17 (1)
18 SEQUENCE FROM N.A.
19 TISSUE=Thyroid;
20 A Isegai T., Oca T., Nishikawa T., Hayaashi K., Otsuki T., Sugiyama T.,
21 A Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
22 A Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
23 A Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
24 A Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
25 "NEDO human cDNA sequencing project.";
26 T Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
27 L EMBL; AK075265; BAC11507.1; --
28 R GO: GO:0016491; F:oxidoreductase activity; IEA.
29 R GO: GO:0006118; P:electron transport; IEA.
30 R InterPro; IPR008151; Phyt_n dehydro.
31 R InterPro; IPR001100; Pyr_redox.
32 R PRINTS; PR00411; PNRDRTASE1.
33 R ProDom; PD139017; Phyt_n dehydro; 1.
34 W Hypothetical protein.
35 Q SEQUENCE 591 AA; 63067 MW; 7029E2119839208E CRC64;

Query Match      22.6%; Score 624.5; DB 4; Length 581;
Best Local Similarity 33.0%; Pred. No. 9.5e-31;
Matches 185; Conservative 81; Mismatches 226; Indels 69; Gaps 18;

14 6 DAVVVGSHNALVSAAVLAREGWSVELENDVLQGAIVTVERFPQYKVDRCSSAHLMR 65
15 36 DAVVIGAGHGLVAVALQELQVNTAVFERHVIGGAATEEIIIGFESRASYLLSLR 95
16 66 HSGIIEBELGANGERYIDCDPNAPAPPAGTDGPG-----IVFHRDLDTACQIERAC 119
17 96 PQ-IYTDLELKGHGLHLENPYSFTPLM--EEGAGSKVPRCLLGTDMAENQKQIAQ-F 151
18 120 GTKDADAYRFAVWWSRSRHVKAFSTPTG-----SNLICAFGLGATARG----- 166
19 152 SQKDAQVFPKYEFPMHLELAIDPLDPAEVDMAAFQHGSLQWRSESLTKELLKAGRI 211
20 167 -NSELRSQF---LAPGDALDDEVDSALKKAALW---FGAGSGPPMPEGTAPNMGAA 219
21 212 LGAQLPRYYEVLTAPIKVLQDQFEEELKATLATAVIGAMTSP--HTPGS---GYVL 265
22 220 LMHVLP-----GRAVGSGLAALASRAVDCATVALGDGVTISIRNSNHW--V 269
23 266 LHHVWGLEGMOGNWYVQGGMGALSDALASSATHTGASIFTEKTVAKVQVNSECCVGV 325
24 270 TTSEGREVHARKVIAGCH-----LITLDLLNGGDFRTDLDHWRKIRVPGIGAVLR 323
25 326 VLEDGTEVRSKWVLSNTSPQITFLKLTQPSWLPBEFLERISQLDTRSPV----- 378
26 324 LATSALPSY-----RGDATTRESTSGQLLVSDRAHLRTAHGALAGELPPRPAVLGM 376
27 379 VAVDELPSFLAAPNAPRGQPLPHHQCS-IHLNCEDTLLHQAFEDAMDGLSPHRFVIEL 437
28 377 SFGSIDTPIAPGHQVTLMSQHQFTRLSGHRDW-ASVAEADRIIVGEMEPAPCFPTDS 435
29 438 IPSSLSDPTLAPPGHVYVSLFTQTFYTLAGGKAWDQERDAYDFVFCIEVYAFGPKDS 497
30 436 VLDRFIQTPRDIESELGMVGNVWVMESLDQMLWRPLPELSGHRVPGADGLYITGAST 495
31 498 VVGRDILTPDLERIPLPGGNIFPHCMSLSDQLYFTRPVPLHSGYRCP-LQGLYLCSGA 556
32 496 HPGGVSGASGRSAAITALSD 516
33 557 HPGGVGGAAGRNAHVAFRD 577

```

RESULT 6
18VYK3

ID	Q8VYK3	PRELIMINARY;	PRT;	556 AA.
AC	Q8VYK3;			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	AT594950/K6M13.10.			
OS	Arabidopsis thaliana (Mouse-ear cress);			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	euroside II; Brassicales; Brassicaceae; Arabidopsis.			
NCBI_TaxID=3702;				
ON	[1]			
RP	SEQUENCE FROM N.A.			
RP	Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,			
RA	Shinn P., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,			
RA	Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,			
RA	Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,			
RA	Nguyen M., Onodera C.S., Palm C.C., Quach H.L., Sakurai T., Satou M.,			
RA	Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,			
RA	Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,			
RA	Scker J.R.;			
RT	"Arabidopsis cDNA clones";			
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,			
RA	Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan Y.W.,			
RA	Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,			
RA	Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,			
RA	Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,			
RA	Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,			
RA	Shinozaki K., Davis R.W., Theologis A., Scker J.R.;			
RT	"Arabidopsis ORF clones";			
RE	Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY070478; AAL49944.1; -			
DR	EMBL; BT005815; AAO64750.1; -			
DR	InterPro; IPR000205; NAD_BS.			
DR	InterPro; IPR008151; Phytin_dehydro.			
DR	ProDom; PD139017; Phytin_dehydro; 1.			
SQ	SEQUENCE 556 AA; 60560 MW; FC861290FD790B5A CRC64;			
	Query Match 20.7%; Score 572.5; DB 10; Length 556;			
	Best Local Similarity 31.1%; Pred. No. 1.5e-214;			
	Matches 191; Conservative 78; Mismatches 217; Indels 109; Gaps 20;			
Qy	6 DAVVGSCHNALVSAAYLAREGHSVELEKDTVLGVANSTVERFPGYKVDGSSAHLMIR 65			
Db	15 EAWTGGGHNGLTAAAYLARGLSVALERRHVIGGAATEEIVFGFPRCSYVQLGLR 74			
Qy	66 HSGIIEELGLGAHGLRVIDCDPNAFAPPAGTGTGPGIVPHRLDLATCSIERACQTDAD 125			
Db	75 PC-IIRELELGHGKLLKSPSPF--TPCLDGRVLLGPDQLNHSEISK-FSKHAD 129			
Qy	126 AYRFVAVWSRSRHVKKAF-----STPTGSLNIGAFGLATARGNSLSRQ----- 173			
Db	130 AYPRY-----EKQLERFCGMDFLLDSTPESLQSASF-----NDKLSNKKMYKAF 176			
Qy	174 -----FLAPGDALLDEYDFSEALKAAALAFCA-OSGPPMSPFG 210			
Db	177 WARCRLQAVSLGHDKYVAFMDLLILASPKNVWFESDVLKSLATDAVIGSTASVHTFG 236			
Qy	211 TAPMVGFALMHVLP-----GRAVGSGLSALASLRNAVDCATVALCDGYSIR 262			
Db	237 S-----GYVLLSHVNGEIDGSKVSWYVEGWSVSNIAANAKEAGASIFITNAEVSFI-- 290			
Qy	263 NSNHWVTTESGREVHARKVIACGHILTLTDLGNCGPDRTTLDHWRRKIRVGPQIGAV- 321			
Db	291 -----LTDSSIVKGVLDAGTRVSSA-ILSNATPTFTVELVPTNLPENFVSAIK 342			
Qy	322 -----LRLATSLPSYRGDATTRSTGGQLQLVSDRAHLTAH-GAALAGE----- 366			
Db	343 NSDYSSATTKNLAVDKLPOFO-CNTNHSGGPPE-----HFCTIHIGAESMDVESA 394			


```
367 -----LPPRAVLGMSF-SCIDPTIAPAGRHQVTLWSOMQPYRLS-GHRDWSVAEAE 417
395 CHDSNGULPSRRPVEMTIPSTLONTISPPCKHVINLFQVTPFKPSDGSWEDPTREAF 454
418 ADRIVGENEAFAPGFTDSVLDRFQTTPREDIESELGMIGGNMVMHVMESLDMMLWRPLPEL 477
455 AQRCPKLDIYAPGSSIIISYDMLTPDLEREIGLTGGINIFHGAMGLDSLFLMRPVKGM 514
478 SCHRVPGADGLYLTCASHTPGGVSAGSRAARIALSD 519
515 SNYRSP-LKGLYLCGSGHPPGGVWAGPGRNAHVQLDLK 555

RESULT 7
FGZ1
Q9FGZ1 PRELIMINARY; PRT; 647 AA.
01-MAR-2001 (TremBLrel. 16, Created)
01-MAR-2001 (TremBLrel. 16, Last sequence update)
01-OCT-2003 (TremBLrel. 25, Last annotation update)
Phytoene dehydrogenase-like.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
STRAIN=Columbia;
MEDLINE=20181125; PubMed=10718197;
Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
Tabata S.,
Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
features of the regions of 3,076,755 bp covered by sixty P1 and TAC
clones."
DNA Res. 7:31-63(2000).
EMBL; AB023033; BAB10768.1; -.
InterPro; IPR000205; NAD_B5.
InterPro; IPR008151; Phytin dehydro.
ProDom; PD139017; Phytin dehydro; 1.
SEQUENCE 647 AA; 705687 MW; 3DC4C5918A8D0FC6 CRC64;

Query Match 20.3%; Score 562.5; DB 10; Length 647;
Best Local Similarity 30.9%; Pred. No. 7.9e-27;
Matches 179; Conservative 78; Mismatches 213; Indels 109; Gaps 20;

6 DAVVVGSHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVFRFPYKVDGSSAHMTR 65
15 DAVVVGSHNGLTAAAYLARGGLSVAVLERRHVIGGAATVEETVPGFKFSRCSYLOGLLR 74
66 HSGIIEELGLGAHLRYIDCDPWAFAPAPGTDGPGIVFHRDLDTACQSIERACGKTDAD 125
75 PC-LIRELEGRHGLKLLKSPSPS---TCLDGRVILLGPOQDLNHSRISK-FSKHDAD 129
126 AYRFVAVWSERSRHMVKAFF-----STPTGSLNLIAGPGLATARGNSLSRQ----- 173
130 AYPYR-----EKOLRFGCFWMDPLDSTTPESLQSASSF-----NDKLSNMKYSAF 176
174 -----FLAPCDALLDEYFSEALKAAALNFGA-QSGPPMSRPG 210
177 WARCIRQAVSLGHKMDVAFMDILLAFASKVINNNWFSDVLKASLATDAVIGSTASVHTPG 236
211 TAPMVGFAALMHVLP-----GRAVGGGALSAAALSRMVAVDGATVALGDGVTISRR 262
237 S-----GVVLLHHVWGTEGKGISVYVEGGMGSVMAIANAKAEAGAEFTNAEVSSEI-- 290
263 NSNFWTTTSGREVRHARKVIAGCHILTTLLDNGGFFRTLLDHRKIRKIVGPGIGAV- 321
291 -----LTDSSSTVKGVLADGTRVRESSA-ILSNATPYRTYVELVPTNVLNPFVSAIK 342
322 -----LRLATSPALSVRGDATTRESTSGLLQLLVSDRAHLRTAH-GAALAGE---- 366
343 NSDYSSATTINLAVDKLPQFQ-CCNINHSGPCPE-----HFGTIHIGASSMDVEHSA 394
```

RESULT 8

```
Q98FP6 PRELIMINARY; PRT; 521 AA.
AC Q98FP6
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN MLR3676
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAPEF303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003002; BAB50521.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR000205; NAD_B5.
DR InterPro; IPR008151; Phytin dehydro.
DR InterPro; IPR003042; Rgm_aminooxygenase.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00420; RGMNOXGNASE.
DR ProDom; PD139017; Phytin dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 521 AA; 54661 MW; 12846C9155B8197 CRC64;

Query Match 20.1%; Score 557.5; DB 16; Length 521;
Best Local Similarity 34.2%; Pred. No. 1.2e-26;
Matches 188; Conservative 75; Mismatches 213; Indels 73; Gaps 28;

QY 1 MSATLDAVVGSHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVFRFPYKVDGSSA 60
DB 1 MTSP-DAVIGGSHNGVAAATLAKAGKVLVLEASVEVGGAARTBEFAPGFRV--SSIA 57
QY 61 HLMTR-HSGIIEELGLGAHLRYIDCDPWAFAPAP-GTDPGIVFHRDLDTACQSIERA 118
DB 58 HLLNRLHPDVVKVTKLELTHGLQPARAD---FLPSAALSCKDGPALMLH-----GA 103
QY 119 CGTKDADAYRRFVAVNSE-----RSHVMKAPST--PP--TGSNLI--GAFG--GLATA 164
DB 104 YGEVLTCASFSSEGSAMKDLRAQLLRVAGILKPLFTRPPDLAGNSLMTASLQGTALAK 163
QY 165 RGNSELQRQFLA-----PGDALLDEYFDSALKAAALAW----FQAQSGP--PMSEPG---- 210
DB 164 KLGGEDVRDFLRVLNKNVAD--LLDEQLRDDRLLKGLLAFOATLGLSHLGPSPSTSLGLLYR 222
QY 211 TAPMVGFAALMHVLPGRVAVGGGALSAAALSRMVAVDGATVALGDGVTISIRNSNMT-V 269
```

```
b 223 LAGEAGGAAGVLPQ-----CGMGAWAAIRAAAEKAGVTIRTSVAAAKIIVEKGHAVGV 278
y 270 TTESGREHARVIAGCHILTT-LDLLG-----NGGFDRTTLDHWRKIRKIRVGGIGAVLRL 324
b 279 TLGTGELFARTVWSAINPATTFLDLVGPREDITGFVRKV-----KTRM-KGDAAKLHL 332
y 325 ATSAIPSYRG-DATTRESTSGQLLVSDRAHLTAHGAALAGELPPRPVAVLGMFSGI-D 382
c 333 ALDRSQFTGADAAAHK-----GRVLAPDPDHVERAFNECKYGFSPBP-VLEITLPSLAD 388
y 383 PTIAPAGRGQVTLMSQWQPYRLSGHRDWSVAEABADRIVGEMEFAPGFTDQSVLDRFIQ 442
b 389 PSIALPDGCVLSAVVQYAFYALK--EGWAAGKPKFIKAWAQLAEAVAGICKSVVHAEILL 446
y 443 TPRDIESEIGYIGNVWVHMSLDQMLWRPLPELSCHRVPGADGLYLTGASTHVGCGVS 502
b 447 TPADIETRYNPGGHHGELQDQMLSRPVGWSGYDTP-VEGLFLAGAGSHPGGGSVS 505
y 503 GASGRSAAR 511
c 506 GAFGLNAAR 514

RESULT 9
9A3T1
c Q9A3T1 PRELIMINARY; PRT; 543 AA.
t 01-JUN-2001 (TrEMBLrel. 17, Created)
t 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
t 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
e Phytoene dehydrogenase-related protein.
n CC3121.
s Caulobacter crescentus.
c Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
c Caulobacteraceae; Caulobacter.
x NCBI_TaxID=155892;
n [1]
p SEQUENCE FROM N.A.
c STRAIN=ATCC 19089 / CB15;
c MEDLINE=21173698; PubMed=11259647;
a Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
a Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
a Potocka I., Nelson W.C., Newton A.C., Stephens C., Phadke N.D., Ely B.,
a DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
a Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
a Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
a Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
t "Complete genome sequence of Caulobacter crescentus.";
l Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
r EMBL; AB005975; RAK25083.1; -.
r PIR; G87635; G87635.
r TIGR; CC3121; -.
r InterPro; IPR008151; Phyt_n_dehydro.
r ProDom; PD139017; Phyt_n_dehydro; 1.
w Complete proteome.
c SEQUENCE 543 AA; 57881 MW; 3FE9EB19403B1DE CRC64;

Query Match 19.5%; Score 538.5; DB 16; Length 543;
Best Local Similarity 33.1%; Pred. No. 1.9e-25;
Matches 180; Conservative 76; Mismatches 234; Indels 53; Gaps 21;

y 6 DAVVCSGHNALVSAAVLAEGMSVEVLEKDTVLGGAVSTVERFPYKVDGSSAHLMI 65
b 8 DAVIIGGHNGLVCAYLAKAGLVKTVCEARGVVGGAATVEFFHGFGR-NSVASTVSL 66
y 66 HSGIIBELGLGAGHLYIDCDPWAFAPAGTDPGPIVFHRLDTCQSIERACTKQAD 125
b 67 NRPVADMGRLGELGTLFRLPSINFLPI---SDDKYIKLGGGLRT-QBSEFKYSRDAE 122
y 126 AYRFRVAVMSERHVMK--ASTPTT-GSNLIGAPGLATARGNSLSRQ-----F 174
b 123 VLPAYAMLDLIG-DILRLAQSTPPNLDGLPGLLRLAURQGRGLAFLSRQKRLDLDF 181
```

```
Qy 175 LAPGDALLDEYDSEALKAALANFCAQSPMSRPGTAPMVGFPAALMHLVLP----- 226
Db 182 TKSARDVLDGWFESFVRAAFG-FDAVVG-NPASDT-PGSAIVLLHHTFGEVNGKKAGW 238
Qy 227 GRAVGGSGALSAAALSARMAVDGATVALCDGVTISIRNSNHWI-VTTESGREVHAKVJAG 285
Db 239 GHAVGGGAIQTQAKAKACEAAGVEILLDAPVEAVHIDGKKAAGVQLVDGRQIMAPIVAN 298
Qy 286 CH-ILTTLLNGGDFD---RTILDHWRKIRVPGIGAV-LRLATSALPSYR---GDAT 337
Db 299 VNPALLYKKLVPPSALTDPFRKAVDGYKN-----GSGTFKNVALSELPSFTCLPGKET 352
Qy 338 TRETSYSGQLLVY----DRAHLTAHGAALAGELPPRPVAVLGMFSGIDPTIAPAGRHV 393
Db 353 AEHQSGGIIVASLDYMDAAY-RDAKGGQIS----KAPIVEMLIPLSSLDTSIAPGQHVA 407
Qy 394 TLWSQWQPTRLSGHRDWSVAEABADRIVGEMEFAPGFTDQSVLDRFIQTPRDLIESELGM 453
Db 408 SLFCQQFAPELPDGRSWDDAREAADLIIDTVQWAPGFKASVILGRMTLSPLDLERKFL 467
Qy 454 ICGNVMEVMSLDQMLWRPLPELSCHRVPGADGLYLTGASTHVGCGVSAGSRAARIA 513
Db 468 IGGDINHGCHSLDQLWATEPLIGHASHRAPIA-GLYMOGAGTHPGGVSGNPGRNAAREI 526
Qy 514 LSD 516
Db 527 LRD 529
```

RESULT 10

```
Q98FP7
ID Q98FP7 PRELIMINARY; PRT; 539 AA.
AC Q98FP7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN MUR3674.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFP303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kareko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
t "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003002; BAB50520.1; -.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 59427 MW; 9DD78A7115A50BED CRC64;
```

Query Match 19.4%; Score 537.5; DB 16; Length 539;
Best Local Similarity 30.4%; Pred. No. 2.2e-25;
Matches 168; Conservative 92; Mismatches 218; Indels 75; Gaps 21;

```
Qy 6 DAVVCSGHNALVSAAVLAEGMSVEVLEKDTVLGGAVSTVERFPYKVDGSSAHLMI 65
Db 5 DAVIIGGHNGLVCAYLAKAGLVKTVCEARGVVGGAATVEFFHGFGR-NSVASTVSL 64
Qy 66 HSGIIBELGLGAGHLYIDCDPWAFAPAGTDPGPIVFHRLDTCQSIERACTKQAD 125
Db 65 PE-IMEDLELPRFLQVISYEGAVFT----RDGYLANRYRDHARREFAR-FSRDAE 118
```

```
/ 126 AYRRFVAVWSRSRHMVK-APSTPTGNSL-----IG-----AFGLATARGNSLSRQ 173
/ 119 AYDRYARDVTRQCFIQPLLMRTAPDPTSFPRDITGELLYLKPKFAGLSAEEMALTLRFW 178
/ 174 FLAPGDALLDEYFDSSEALKAAALW---FGAQSQPPMSEPGTAPWVGFPAALMHVLP--- 226
/ 179 TWSISD-FLDEYFEDVDVVKANFALSGIIGTALGP--MSPGTA-----VLLHHYMGVEVDS 231
/ 227 -----GRAVGGSGALSAALASRAMVDGATVAG---DGVTSIRRNNSHWTVTTESGREVHA 279
/ 232 VGAWYARGRGCAVTKAALASGKASGGTIRTGAEVDHVLVSRGKAK--GVVLAGEBEVVG 289
/ 280 KXVLAGCHILTTLLDGLNGGDFRITLTHWR-----KIRVPGPGCAVLRATATSLPS--- 331
/ 290 KLVVSNADVKTFLKVB---EKELPDIFLRVNVFKIR---GSSGKVNIALDLSLPEPPA 343
/ 332 -----YGDATTRESTSGIQLVSDRAHLRTAHGAALAGELPPRAVLGMSFGSIDPT 384
/ 344 LAKDSPVYRGDMHPTDSIERMERAYDDWK-----AGWSADPFLDWVPTTLTPT 393
/ 385 LAPAGRHQVILWSQWQVRLSGHROWASV-AEADRIVGEAEAFAPGFTSDVLDRIQT 443
/ 394 MAPFGKHMSFCVQVAPPKVG--RDWTDADGFAESVVAQIAEYSPGFRDRIVHMEVRT 452
/ 444 PRDIESEIGX:GGNVHVMESLDQMLMRPLPELSGHRVPGADGLYLTGASTHPGGGVSG 503
/ 453 PREIEAEVLTEGNIFOGELTFDQLFNRPVPGVYQVRSR-VGGLYVCGSSTHPGGGVWG 511
/ 504 ASGRSAARIALSD 516
/ 512 APGRNAAAEILRD 524
```

RESULT 11

/W24

Q974W1

ID

AC

Q974W1

DT

01-DEC-2001

(TRENBLrel. 18, Created)

01-OCT-2001

(TRENBLrel. 18, Last sequence update)

01-OCT-2003

(TRENBLrel. 25, Last annotation update)

Phytoene dehydrogenase related protein.

SS02422

Sulfolobus solfataricus.

Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

Sulfolobus.

NCBI_TaxID=2287;

[1]

SEQUENCE FROM N.A.

STRAIN=ATCC 35092 / DSM 1617 / P2;

MEDLINE=21332296; PubMed=11427726;

She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,

De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,

Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;

"The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";

Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

EMBL; AE006842; AAK62556.1; ..

PIR; G90413; G90413.

GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.

GO; GO:0006118; P:electron transport; IEA.

InterPro; IPR000759; Adrnxd_reductase.

InterPro; IPR001327; FAD_Pyr_redox.

InterPro; IPR000205; NAD_BS.

InterPro; IPR008151; NAD dehydro.

InterPro; IPR001100; Pyr_redox.

PRINTS; PR00419; ADXRDATASE.

PRINTS; PR00368; PADPNA.

PRINTS; PR00411; FNRDRTASII.

ProDom; PD139017; Phyt_n_dehydro; 1.

Complete proteome.

```
SQ SEQUENCE 518 AA: 57944 MW; B538CF7B74ACD72B CRC64;
Query Match 19.3%; Score 533; DB 17; Length 518;
Best Local Similarity 30.9%; Pred No. 4e-25;
Matches 168; Conservative 93; Mismatches 226; Indels 56; Gaps 21;
QY 5 LDVVVGGSHALVSAAYLAREGMSVEVLEKDTVLGGAVSTVERPPGKYVDGRGSAHLMI 64
DB 2 IDVALIGGSHGLVTAAYLAKAGLKVAVFERREIVGGASVTEELWPSIKVSTGAYVLSLL 61
QY 65 RBSGIIIEELGAGHLRYIDCDPMAPAPGTPGPGIVPHERDLQATQOSLERACGTDA 124
DB 62 RPK-IIEELKREPGKLYLKDPLF---LPFENGKLYIWSSELEKTKETEK-FSKNDA 116
QY 125 DAYRRFVAVW---SERSRHVMKAFSTPP---TGSNLICAPGLATARGNSLS-----RQF 174
DB 117 KNYKKWKFVDFLFAEMADFFW--LNPPPHIDEASNLINIPRG---NVNBEALSFLRTF 170
QY 175 LAPGDALLDEYFDSSEALKAAAL---WFGAQSQPPMSEPGTAPWVGFPAALMHVLP--- 226
DB 171 MODAKSLDEYFETDEVKSALAEADVVGTFASP--STPGTA-----VVLAAHVIGEVNGVK 224
QY 227 ---GRAVGGSGALSAAL---ASRMADVGTVALGDGVTISRNSNHWTVTTESGREVHA 280
DB 225 GAWGVVEGGMSEVTQALAKRSAEHLGVEIYTNAEDEV--LVKNGRVEGKILKNGKTINAK 282
QY 281 KVIACCHILTT-LDLLNGGDFRITLTHWRKIRVPGIGAVLRLA--TSALPSYRGDAT 337
DB 283 IIVSNADPKTTFFKLLRNALEEDFL---RRVALKNVGVSVFXIVGYLEELPDFGNGKS 338
QY 338 TRESTSGIQLVSDRAHLRTAHGAALAGELPPRAVLGMSFGSIDPTIAPAGRHQVTLWS 397
DB 339 LSPHIASELIMPNEVIEKAYDDARALGYSRPWLINIQSSVDPTVAPPGKPSIFIG 398
QY 398 QMQPYRUGSHDNASVAEAEADRIVGEAEAFAPGFTSDVLDRIQTPTDIESELMGTGN 457
DB 399 QYLVYD-SKND--EMKEIAETFEKIKEPAPNFKIKYE--VLTELDIERRERGINGN 453
QY 458 VMHVMSLDQMLMRPLPELSGHRVPGADGLYLTGASTHPGGGVSGASGSAARIALSDS 517
DB 454 IFHLDMPDQLYFFRPLIGYSIDYTP-IRGLYLCGSGTHPGGVGAGPNAHVKILEDL 512
QY 518 RRG 520
DB 513 RKG 515
```

RESULT 12

Q974W1

ID

Q974W1

DT

01-DEC-2001

(TRENBLrel. 19, Created)

DT

01-DEC-2001

(TRENBLrel. 19, Last sequence update)

DT

01-OCT-2003

(TRENBLrel. 25, Last annotation update)

DE

Hypothetical protein ST0549.

GN

ST0549.

OS

Sulfolobus tokodaii.

OC

Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC

Sulfolobus.

OX

NCBI_TaxID=111955;

RN

[1]

SEQUENCE FROM N.A.

RC

STRAIN=JCM 10545 / 7;

RX

MEDLINE=21456156; PubMed=11572479;

RA

Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,

Sakine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,

Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,

Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,

Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Yamagishi A.,

Oshima T., Kikuchi H.;

RT

"Complete genome sequence of an aerobic thermoacidophilic

Crenarchaeon, Sulfolobus tokodaii strain7.";

RL

DNA Res. 8:123-140(2001).

```
DR EMBL; AP000982; BAB5546.1; --
DR GO: 0006118; P: electron transport; IEA.
DR InterPro; IPR000759; Arndx_reductase.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR InterPro; IPR002005; Rab_GDI_REP.
DR PRINTS; PR00419; ADXRD7ASE.
DR PRINTS; PR00891; RABGDIREP.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 517 AA; 57965 MW; E129448C73A2A24 CRC64;

Query Match      19.0%; Score 525; DB 17; Length 517;
Best Local Similarity 30.1%; Pred. No. 1.3e-24;
Matches 165; Conservative 89; Mismatches 228; Indels 66; Gaps 20;

Xy 1 MSAFDVAVVGSGNALVSAAYLAREGMSVEVLEKDTVLGCAVSTVERPPGYKVDGSSA 60
Db 1 MNRIVDVVIGGNGHGVVAAYLAKGLKAVTERNVIGGATVTELPFGIKVSTASTV 60
Xy 61 HLMTRHSGIIIEELGLGAGRLYIDCDPWAFAPAGTGDGFGIVFHRDLDTQCSIRACG 120
Db 61 FSLPRKE-IIDDLRLYDFGLKVLKDPSPVPPG---QNKSIITMTSTKTKVKEIK-PS 115
Xy 121 TKDADAYRRFV---AVWSERSRHVKAFSTP---PTGSLNIGAFGLATARGNS-ELSRQ 173
Db 116 KDAQSTKFWKLEIFSEIDFLI--LNKPINLSABELLKFLGLNVDENVTALTIARM 173
Xy 174 FLAPGDALLDEYDFDSEALKAAAL---AMFGAQSGPPMSEPGTAPMVGFAALMHVLP 226
Db 174 FPDGKSFLEDFESESVEKALIEDSVGTYASP--STPGTA---YVLLHNFGEVNGV 227
Xy 227 ---GRAVGGSGALSAALARMADGATVALGCVTS--IRNSNHVTVTESGREVHARK 281
Db 228 KGAWYVEGGMGVNSALRRVABYGAIFLNSEVDEIIVKNGEAKGIKRGKIESKI 287
Xy 282 VIAGCHILT-LDLLGNGGPDRTTLDHWRKIRVGPICIGAVLR--ATSALPSYRGDATT 338
Db 288 VLSNADPKTFLKLRNAELD---EPIRKVNAKTIGVSPKINGYIEELDFGNGKSL 343
Xy 339 RESTSGQLLVSDRAHURTAHGA---LAGELPAPRVLMGMSFGIDPTIAPA 388
Db 344 RPEHASELIIPSVSVEKAYLDKAKLVGSRPWLINIP-----STVDPTLAPQ 393
Xy 389 GRHQVTLMSQVPLSGHSDWASVAEADRTVGEAFAPGFTDVLDRFIQTPDIE 448
Db 394 GKPVTFIFGVYVY---SKDLKIKDKIAISIEKVFAPNFPIKYE--ILPLDIE 447
Xy 449 SELGMIGNVMHVMESLDQMLRPLPELSGHRVPGADGLYLTGASTHPGGVSGASGRS 508
Db 448 RRFQIMEGNIFHLDMTDQQLYFFRPLIGYSNRYTP-IKNLYLCGSGTHPGGVTGAPGN 506
Xy 509 AARIALSD 516
Db 507 ASMEILKD 514

RESULT 13
ID Q89FW3 PRELIMINARY; PRT; 539 AA.
C Q89J01
T 01-OCT-2002 (TrEMBLrel. 22, Created)
T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E Phytoene dehydrogenase.
N XAC2744.
S Xanthomonas axonopodis (pv. citri).
C Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
C Xanthomonadaceae; Xanthomonas.
X NCBI_TaxID=52829;
P SEQUENCE FROM N.A.
C STRAIN=306 / ATCC 13902 / XV 101;
```

```
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quesgio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo E.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chabergo J., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fortighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.S.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.P.,
RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
RA Setubal J.C., Kitajima J.B.;
RA --Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.;
RL Nature 417:459-463(2002).
DR EMBL; AB011914; AAM37589.1; --
DR InterPro; IPR008151; Phyt_n_dehydro.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 57459 MW; 5E9801191CB48C3D CRC64;

Query Match      18.4%; Score 509; DB 16; Length 539;
Best Local Similarity 32.3%; Pred. No. 1.3e-23;
Matches 182; Conservative 73; Mismatches 213; Indels 96; Gaps 26;

Qy 5 LDVAVGSGNALVSAAYLAREGMSVEVLEKDTVLGCAVSTVERPPGYKVDGSSAHLMI 64
Db 16 LDALIIGAGHNGVCAAYLAKRGRVLVLEAREVVGGAATEEFHGFR-NSVAAYTVSL 74
Qy 65 RHSGIIEELGLGAGRLYIDCDPWAFAP-----PAGTD-----QFGIVFHRDLDTQ 113
Db 75 LQPRVIADLALERHGRV-----PRINNPLPNCGYLLAGAG-----RTAQ 118
Qy 114 SIERACTKDADAYRRFVAVWSERSRHVKAFSTPP-----TGSNLIGAFGLATA----- 164
Db 119 DVAR-PSARDAALPAEYA-RLEQLADVLRLALALQPPDVTGGWLQALPOLWRACKLGL 176
Qy 165 --RG-----NSLSRQFLAPGDALDEYDFDSEALKAAALAW---FGAQSGPPMSEPGTAPM 214
Db 177 QLQGLPLTLRQELLDLTISAAEVLDRWFESAPIKALFGFDGIVGNVYASP--HTPGTA-- 232
Qy 215 VGFPAALMHVLP-----GRAVGGSGALSAALARMADGATVALGCVTSIRNSNH 266
Db 233 --YVLLHHVFGQSGNGYKAWGHAIGMGAITQMAASAREYGAQLTGCACVERVLVEQGH 290
Qy 267 WT-VTTESGREVHARKVIAGCH-ILTTDLGNGGPDRTTLDHWRKI---RVGPGIGAV 321
Db 291 AVGVVITAGETVBARAVANVNPVKLYQLMRPDDVPATT-----REIAHYRCGSGTFR- 345
Qy 322 LRLATGALPSY-----RGDAITRESTSGQLLV-----DRAHLRTAHGAALAGELPPRPA 372
Db 346 MNVALSELDFALPQNGD-----HLTAGIILAPSLDYMDRAW-----HDAFAFG-WSRSEV 396
Qy 373 VLGMSPSGIDPTIAPAGRHQVTLMSQVPLSGHSDWASVAEADRTVGEAFAPGFG 432
Db 397 VEILLIPSTLDDTLAPGQHVASFQCHVAPQLPDGRRDDHREVDLMIATVERVAFQF 456
Qy 433 TDSVLDRIQTTPDIESELGMIGNVMHVMESLDQMLRPLPELSGHR--VFGADGLYL 490
Db 457 AASVLGRQVLSPLDLERMFGLVGGDIFHGALSLNQLFSGARPLGQGVARGAVP---GLYL 513
Qy 491 TGASTHPGGVSGASGRSAARIAL 514
Db 514 CGSGTHPGGVTGAPGHNAARVIL 537

RESULT 14
ID Q89FW3 PRELIMINARY; PRT; 545 AA.
C Q89FW3
```

Q39FW3; 01-JUN-2003 (TReMBLrel. 24, Created)
01-JUN-2003 (TReMBLrel. 24, Last sequence update)
01-OCT-2003 (TReMBLrel. 25, Last annotation update)
BLR6586 protein.
Bradyrhizobium japonicum.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
NCBI_TaxID=375;
[1]
SEQUENCE FROM N.A.
STRAIN=USDA 110;
MEDLINE=2248498; PubMed=12597275;
Kaneke T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
Kohara M., Matsumoto M., Shimo S., Tsurucka H., Wada T., Yamada M.,
Tabata S.;
"Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
DNA Res. 9:189-197(2002).
EMBL; AF005959; BAC51851.1; -.
GO: 0016491; F:oxidoreductase activity; IEA.
GO: 0006118; P:electron transport; IEA.
InterPro: IPR000205; NAD BS.
PRINTS: PR000103; Pyridine_redox_2.
Complete proteome. PNDRTASEII.
SEQUENCE 545 AA; 58480 MW; 0B2DDC6F39C0B783 CRC64;
Query Match 18.0%; Score 497; DB 16; Length 545;
Best Local Similarity 30.5%; Pred. No. 7.4e-23;
Matches 174; Conservative 77; Mismatches 217; Indels 102; Gaps 25;
5 DAVVVGSHNALVSAYLAREGSHVEVLEKDTVLGGAVSTVERFGYKVGKSGSAHMR 65
14 DVVIIGAGHNGTCAAYLAWAGLRVVRVERKVGWGAATVEEPHGFPR-NSVAYTVSL 72
66 HSGIIEELGLAGHLYIDCPWAPAPPTG---DGPGLVPHRDLDATCQSIERAGTK 122
73 NPQVIRDLGLAQGLRIVERAQNPLPADGSYLLTGE---RTKASVAR-LSAH 123
123 DADAYRRFVAVWSESRH-----VMKAP--STPTGSLNIGAFG----- 159
124 DADALDGE-----SRELEDIADVLQFVLRAPP---NLLDGFQTNAREAVNALQSAN 173
160 ---GLATARGNSLSRQFLAPGDALDEYFDESKAALAWFGAQSGPPMS--EPGTAPM 214
174 ILRGITLQSRSLD-LFTRSGEMLDERFEDLVK-ALPGFDIVGNYPYAGSA-- 229
215 VGFALMHVLP-----GRAVGGSGALSAAALASRMAYDVGATVALGDGVT-IRNSN 265
230 ---YVMLHAFGEVANGKGVWGHGALGNGAITQAWABAAQGRGVAIDTDAGVREIVR 287
266 HWTVTESGREVHARKVTAGCH-----ILTTDLGNGGDFRTLLDHWKRTVRPGIG 319
288 AVGVLENGATIRAKYVAANVPKLLVTRFAADALPKDFLAR--IRHWKN-----GSG 339
320 AV-LRLATSALPSYR-----GDATTRESTSLQLLV-----DRAHL-RTAHGAALAGELP 368
340 TFRMVALDLRPSFTALPDGCD-----HLTSGIILAPSLGYMDRAFLDARAGHS----- 389
369 PRPAVLGMSFGIDPTAPAGRHQVTLWSQVRLSGHRDNASVAEADRVIGMEAP 428
390 REPVEVLEIPSTLDDTLAPGKKHVASLFCOHVAPELPDSRSDDHDEVDLMIATVDSY 449
429 APGFTDSVLRFIOTPRDISELGICGNVHVMESLDQMLWRPELSGHRVPGADGL 488
450 APGASSVLGQILSPDLREQFLLGSDIFHGALTINLQFSARPMLGADYRGP-LKGL 508
489 YLTGASTHPGGVSGASGRSAARIALSDSR 518
509 YHCGSGAFPGGVGTGAPG-ENAAQAAILRDHR 538

RESULT 15

Q98BS8 PRELIMINARY; PRT; 524 AA.
ID Q98BS8
AC Q98BS8;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Probable dehydrogenase.
GN ML15443.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFE303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneke T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
Takeuchi C., Yamada M., Tabata S.;
"Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RNA Res. 7:331-338(2000).
RL EMBL; AF003006; BABS1894.1; -.
DR InterPro: IPR008151; Phytin dehydro.
DR ProDom: PD139017; Phytin dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 524 AA; 55519 MW; 65B9A7400C90B282 CRC64;
Query Match 16.9%; Score 469; DB 16; Length 524;
Best Local Similarity 29.8%; Pred. No. 3.9e-21;
Matches 161; Conservative 81; Mismatches 238; Indels 60; Gaps 22;
QY 1 MSARFADAVVVGSHNALVSAYLAREGSHVEVLEKDTVLGGAVSTVE-RFGYKVDKSS 59
Db 1 MSERF-DALFVGAGHNSLCAALALKGWKTGIFERSSTIGGAVQTRETLPGFHRDFGM 59
QY 60 AHLMRHSIIIEELG--LGAHGLRY---IDCDPWAFAPAPGTGPGIVFHRDLDATCQS 114
Db 60 NLSLFAGSAFHRKYANELKTQGLEFAPVADC---FASAPP--DGRWFGVNDLEKTASR 113
QY 115 TERAGCTDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGGLATARGNS---ELS 171
Db 114 M-AAFSAADAATWRKLVAAFPGEAEHLFRLIGSPMSAFALAGTAWNLWRKKGAVGALDTG 172
QY 172 RQFLAPGDALLDEYFDESKAALAWFGAQSGPPMSBEGTAPMVGFALMHVLP----- 226
Db 173 RLLSSPRTWLEETFESEPYRATLATWGMH-----DFAPDIAGGAVFPYLESMANQS 225
QY 227 -GRAVGGSGA--LSAALASRMAYDVGATVALGDGVTISIRNSNHMTVTTSBGRVH--ARK 281
Db 226 FQWVLGKGADTIIRALAGMVTSGAGKIVTGAESVETVSNKATGVRLTSGETHATKA 285
QY 282 VIAGCHILTTLDL-----GNGGDFRTLLDHWKRTVGVGCI GAVLRLATSALPSYRGA 336
Db 286 VIAGVAPKALTGTLPGSGNGAGFDTA-----MKFYAPGT-MWHLALDELPGWRAGS 339
QY 337 TTRSTSGQLLVSDRAHLRTAHGAALAGELPPRPAYLGMFSFGIDPTIAPAGRHQVTLW 396
Db 340 ELRQ-FAYVTLSPSLDAMSR-TYQAMAGMLPDSFVLVVGQPTAIDPSRAPQGH--VLW 395
QY 397 SQWQ--PYRLSGH-----RDWASVAEADRVIGMEAFAPGFTDVLDRFIOTPRDI 447
Db 396 VQWRMLPAETIGDAGGKIAPAHMDQVADYAEVRLDIETIYAPGLSKILGRSVFSIDL 455
QY 448 ESE-LGMIGGNVHVMESLDQMLWRPELSGHRVPGADGLYLTGASTHPGGVSGASG 506
Db 456 ERENPNLVGGDQVCGSHLQNFILFRPARGVAGWNP-VGNLHLTGAATWPGAGTGAASG 514

Search completed: February 29, 2004, 14:51:13
Job time : 48.746 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:27:18 ; Search time 8.96629 Seconds
(without alignments)
3089.496 Million cell updates/sec

file: US-09-941-947A-38

Effect score: 2768

Sequence: 1 MSALFDDVVGSGHNLVSA.....ALSDSRGKASQWNRSSRS 532

Scoring table:

Gapop 10.0 , Gapext 0.5

Marched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327	11.8	469	P49_STRLI	P06108 streptomyc
2	303	10.9	535	Y997_MYCTU	Q10555 mycobacteri
3	285	10.3	508	CRTI_STRSE	P54971 streptomyc
4	268	9.7	507	CRTI_STRGR	Q01671 rhodobacter
5	226	8.2	495	CRTD_RHOSH	P48537 cercospora
6	224	8.1	621	CRTI_CERNC	P21334 neurospora
7	208	7.5	595	CRTI_NEUCR	P54980 rhodobacter
8	203	7.4	518	CRTI_RHOSH	P21685 pantoea ana
9	191	6.9	492	CRTI_PANAN	Q02861 myxococcus
10	189	6.8	529	CRTI_MYXXA	P22871 erwinia her
11	179	6.5	492	CRTI_ERWHE	P17059 rhodobacter
12	176	6.4	494	CRTD_RHOC	P54982 phycomyces
13	176	6.4	583	CRTI_PHYBL	P54978 agrobacteri
14	174	6.3	501	CRTI_AGRAU	P17054 rhodobacter
15	174	6.3	524	CRTI_RHOC	P54979 myxococcus
16	158	5.7	517	CRTJ_MYXXA	P56601 myxococcus
17	135	5.0	471	PPOX_MYXXA	Q9hvf1 pseudomonas
18	134	4.8	507	MQO2_PSEAE	Q24163 nicotiana t
19	131	4.8	548	PPOC_TORAC	Q55629 synchocyst
20	130	4.7	471	PPOX_MYCTU	Q53303 mycobacteri
21	129	4.7	452	Y782_SXYN3	Q04829 halobacteri
22	129	4.7	474	DLDH_HALUO	Q05355 streptomyc
23	128	4.6	555	HYDL_STRHA	O50311 chlorobium
24	127	4.6	469	DLDH_CHLVI	P12255 bordetella
25	127	4.6	3590	PHAB_BORPE	Q94526 drosophila
26	125	4.5	1001	ORKI_DROME	Q32434 propionibac
27	124	4.5	527	PPOX_PROFR	P55349 rhizobium s
28	121	4.4	417	Y4AB_RHLSN	P19471 streptomyc
29	119	4.3	529	SP15_STRGR	Q9r008 mus musculu
30	119	4.3	395	KIME_MOUSE	Q96tq9 homo sapien
31	119	4.3	567	FIG1_HUMAN	P98161 homo sapien
32	118	4.3	4303	PKD1_HUMAN	O14686 homo sapien
33	118	4.3	5262	MLL2_HUMAN	

RESULT 1

ID	P49_STRLI	STANDARD	PRT	469 AA
AC	P06108			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	P49 protein.			
GN	P49.			
OS	Streptomycetes lividans.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomycetes.			
OX	NCBI_TaxID=1916;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=66 / 1326;			
RX	MEDLINE=87231086; PubMed=3453116;			
RA	Burnett W.V., Henner J., Eckhardt T.;			
RT	"The nucleotide sequence of the gene coding for XP55, a major secreted protein from Streptomycetes lividans.";			
RL	Nucleic Acids Res. 15:3926-3926(1987).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			

CC	EMBL; Y00142; CAA68336.1; -.
DR	PIR; S09189; S09189.
DR	InterPro; IPR000759; Adrindx_reductase.
DR	InterPro; IPR000205; NAD_BS.
DR	InterPro; IPR008151; Phytin_dehydro.
DR	PRINTS; PR00419; ADXRDTASE.
DR	ProDom; PD139017; Phytin_dehydro; 1.
SQ	SEQUENCE 469 AA; 49740 MW; C2AE7533C7C701CB CRC64;

Query Match	11.8%; Score 327; DB 1; Length 469;
Best Local Similarity	27.6%; Pred. No. 8.6e-15;
Matches 147; Conservative	64; Mismatches 223; Indels 98; Gaps 25;
QY	5 LDVVVSGSHNLVSAAYLARBSVSEVLEKOTVGGAVSTVE-RFPGYKVDGSSAHLM 63
DB	2 LDVVVGGAGPNGLTAAYELARRGFVAVFEAQGTGGGARTTELTLPGRPHDPCSAHPL 61
QY	64 IRHSIIIEELGLAHGRLYICDPWAFAPAGTDPGIVFHRDLDTATCOSIRACCTKD 123
DB	62 GINPAPRGRPLERYGLENLH-PGLPMAHFP--DGAVALSRVGETAASF---GARD 114
QY	124 ADAYR-----RFVAVWSRSRHVAFSTPTPTGNLIGAFGLATARGNSLSQFLA--P 177
DB	115 AGPYRRLIERFLPRWDTLARDFM---SLPLT-----ALPRDPVTLARFGLVGLP 160
QY	178 GDALLDFYDFSEALKKALAWFGAQSGPPNPEPTAPWVGFAALMHVLPFGRAVGGSGALS 237


```

161 PSTWLMERFDEKAKTLEAGLVAVH-----MAPLGGFAT-----GALGVFALA 204
238 AALASMAVVGATVAGLDGVTSTRNSNHWTTTSGRSHVARKVI-----ACCHIL 289
205 AHARGWVARGGSSQSDAATVYK-----DUGGAVHTDYEVRKLDLPPARAYVL 255
290 TT-----LALLGNGGDFRTTLDHWRKIRVGPQIGAVLRATLSALPSYRGDAATTRESTS- 343
256 DTSPALTARIALGLG-----SHY-ANYRGPSPVKIDVALDQVP-----WTAEPERSA 302
344 GLQLLVSDRAHLSTA-HGAALAGELPPRAVLGMSFGSDIPTIAPAGRHQVTLNSQWQPY 402
303 GTVOIGADSTEIGAAHLHAPSGTDPRAPERFLITVQPSVADPTAPAGRH--VFWAY----- 356
403 RLSCH--RDWASVAEAEADRIVGEMAEAFAGFETSDVLDRTFIQTPRODIES-ELGMITGNNVM 459
357 ---GHVPNGWNGDL---TDAMEQLERFAPGPRDRVLARATAGPAELARANNVYGGDIS 410
460 HVMSLDQMLWRPLBELSGHRVPGADGLYLTCGASTHPGGGVSAGSGSAAR 511
411 SGAVSGLQLLL-RPKISLFPYSTP-RPAVFICSSATPPGPGVHGMSGHNAK 460

```

RESULT 2

```

897 MYCTU
D 1897 MYCTU STANDARD; PRT; 535 AA.
C Q10555;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Hypothetical protein RV0897C/MT0921/MB0921C.
E RV0897C OR MT0921 OR MTC31.25C OR MB0921C.
S Mycobacterium tuberculosis, and
S Mycobacterium bovis.
C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
C Corynebacterineae; Mycobacteriaceae; Mycobacterium.
X NCBI_TaxID=1773, 1765;
N [1]
P SEQUENCE FROM N.A.
C SPECIES=M.tuberculosis; STRAIN=H37RV;
X MEDLINE=98295987; PubMed=9634230;
A Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
A Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekaiia F.,
A Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
A Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
A Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
A Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
A Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
A Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
T "Deciphering the biology of Mycobacterium tuberculosis from the
T complete genome sequence.";
L Nature 393:537-544(1998).
N [2]
P SEQUENCE FROM N.A.
C SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
X MEDLINE=22206494; PubMed=12218036;
A Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
A Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
A Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
A Delcher A., Utterback T., Weidman J., Kucuri H., Gill J., Mikula A.,
A Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
T "Whole-genome comparison of Mycobacterium tuberculosis clinical and
T laboratory strains.";
L J. Bacteriol. 184:5479-5490 (2002).
N [3]
P SEQUENCE FROM N.A.
C SPECIES=M.bovis; STRAIN=AF2122/97;
X MEDLINE=22709107; PubMed=12788972;
A Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,
A Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
A Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
A Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

```

```

RT The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73101; CAA97372.1; -.
CC EXBL; AE006979; AAK45167.1; -.
CC EMBL; BX248337; CAD93782.1; -.
CC PIR; F70782; F70782.
CC TIGR; MT0921; -.
CC
CC TuberculList; RV0897C; -.
CC InterPro; IPR001613; Amineoxid_fl.
CC InterPro; IPR000205; NAD_BS.
CC InterPro; IPR008151; Phyt_n_dehydro.
CC PRINTS; PR00757; AMINEOXDASEF.
CC DR ProDom; PD139017; Phyt_n_dehydro; 1.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 7 27 POTENTIAL.
CC TRANSMEM 509 529 POTENTIAL.
CC SQ SEQUENCE 535 AA; 56233 MW; CDAAG92094DCB7D6 CRC64;
Query Match 10.9%; Score 303; DB 1; Length 535;
Best Local Similarity 24.4%; Pred. No. 4.1e-13;
Matches 147; Conservative 74; Mismatches 214; Indels 168; Gaps 25;
QY 6 DAVVCSGNAVSAYLAREGMSVEVLEKDTVLGAVSTVERFPGYKVDROSSAELM-I 64
DB 9 DVVVGGHNGLVAAAYLARAGLRVLLERLAQTGGAASVIOAFDGEVVAUSRYSL 68
QY 65 RBSGIIIEELGAHGLRYDCDFWAFAPAPGTDG-PGIVFHRDLDTACQSIERACGTD 123
DB 69 LPSRIVADLGAPEVRLAR---RPFSSYTPAPATAGRSGLLIGTGPRAAHLAAGAAPD 124
QY 124 ADAYREF-----VAVW-----SERSH-----VMAFSTPTPTGSL 154
DB 125 AHGFAAFYRRCRLVTARLWPTLIEPLRTREQARDIVYEGHEAAAAAQAQVDEPIGHAI 184
QY 155 IGAFGGLATARGNSLSROFLAPGDALLDEVPSLKAALAWFAQSGPPMSEBGTAPM 214
DB 185 AGAVA-----NDLLRGVIAT-DALIGTF-----ARMHEPSLMQN 217
QY 215 VGFPAALMEVLPGRVAV-----CGSGALSAAASRAVAVDQATVALG-----DGVTSIR 261
DB 218 ICF-LYHLVGGGTGVWHVPIGSGSVTSALATAAAREGAEIVTGADYFALDPOG--TVR 273
QY 262 RNSNHWTVTTESGRE--VHARKVIAGCHILATLIDLNGSGDFDRTTLDHWRKIRVGP 319
DB 274 YESD-----GSDGAELVGRFVTVGTPAVLASLLG-----EPVAALAPGAQ 316
QY 320 AVLRATLSALPSYRGDAAT-RESTSGLQLVSDRAHLRTAHLTAAGALAGELPPRAVLGNSF 378
DB 317 VKVMVVRLEPLRLRDSVTPQAFAGTFHVNETSQDLDAAYSQAASGRLPDPLPCEAYCH 376
QY 379 SGIDPTI-----APAGRHQVTLNSQWQVRLSGHRDVAEAEADRIVGEMEPAPGFT 433
DB 377 SLTDSILSARLDAGAQTHLVFGLHTPSHVPD-----TEGLAERTAAVLA---SL 426
QY 434 DSVLDRFIQ-----TPRDISELGMIGGVNVMVMSLDQMLNRPLPEL 477
DB 427 NSVLAEPIDQVLWTDQASKPCIETTTLTDLQRTKLTGCGNIFHGALS-----W----- 474
QY 478 SGHRVPGADG-----LYLTGASTHPGGVSGASGSAARIALS--D 516
DB 475 -----PFADNDPDLTPARQWGVATDHERIMLCUGSAGRRGAVSGIGGHNAAMAVLACIA 529
QY 517 SRR 519
|||

```


428 -----FPASQSIYGLSPHGALA-----SLKPLARTA-----LPGLWLAGGGAHPGAGV 472
502 --SGASGSAARIALSD 516
473 PMAALSGRAAEAILAD 489

RESULT 6
TI_CERNIC
P48537;
01-FEB-1996 (Rel. 33, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase).
PDB1.

Cercospora nicotianae.
Eukaryota; Fungi; Ascomycota; Pezizomycotina;
Dothideomycetes et Chaetothyriomycetes incertae sedis;
Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cercospora.
NCBI_TaxID=29003;
[1]

SEQUENCE FROM N.A.
STRAIN=ATCC 18366;
MEDLINE=94368091; PubMed=8085820;
Brenshaf M., Daub M.E.;
"Isolation, sequence, and characterization of the Cercospora
nicotianae phytoene desaturase gene";
Appl. Environ. Microbiol. 60:2766-2771(1994).

-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COPACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene desaturase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U03903; AAB86988.1; --
PIR; T48646;
InterPro; IPR008150; Bac phytoene dh.
InterPro; IPR008151; Phyt dehydro.
ProDom; PD139017; Phyt dehydro; 2.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
Transmembrane.
NP BIND 11 44 FAD (ADP PART) (POTENTIAL).
TRANSMEM 536 556 POTENTIAL.
SEQUENCE 621 AA; 69529 MW; 61B9EA7784963CA8 CRC64;

Query Match 8.1%; Score 224; DB 1; Length 621;
Best Local Similarity 22.1%; Pred. No. 9.9e-08;
Matches 129; Conservative 73; Mismatches 245; Indels 136; Gaps 22;

7 AVVVGSHALVGAAYLARGSWVEVLEKDTVLGAVSTVERPPGVKVDKRSASHLMITH 66
10 AIVLGSGVGVSTAAALARGAFVTVLEKNNFTGRCSLIHH-EGYRFPQGSLLLL--- 65

67 SGHIE----ELG--LGAHSLRYIDCDPMFAFPAPGTDGPIVFRHDLDTACQSIERACG 120
66 PGLFHRFTFAELGTSLEQGVKLLKCPNVM---IHFDGKEKFTLSDDLVMKTEVEKWEK 122

121 TKDADATRRVAVVNSRSHVMKAFSTPPFGSLNIGAFGLATARGNSLSQFLA--PG 178
123 ---KEGTRYLEFLKESGHYELS-----VREYLLRNFEGL-TAMLRPEFLHLLQLHPF 173

179 DAL---LDEYFDSEALKAALAWCAQSG-PPMSEPGTAPMWGFAALMHLVPPGRAVGGSG 234

DB 174 ESIWTRACKYFWTERLRVRVFTFGSMYMGSPFDAPGTVSLQLQYTELAEGL--WYPVGGFH 231
QY 235 ALSAALASEMAVDGATVALGCGVTSIRNSNHWI---VTTESGEVHARKVIACHILTT 291
DB 232 RVVEALVXIGREBGVDFMETAVKILLSEDDGVGAKVELEDGRLEADVWNNSDLVYA 291
QY 292 LDLLGNGGDFRTILDHWRKIRVGVGIGAVLRATSALPSYRGSDATTRESTGGLQLVSD 351
DB 292 YEKL-----LPKTPVAESILKGRPGSCSSISFYWALDRQ 325
QY 352 ZAHLETAHGAALAGE-----LPRPRAVLGMSFGIDPTTIAPAGRHQVTLW-- 396
DB 326 VPELF-AHNIFLADEYRESFDSIFKGLHLPDEPSFYVNVPSRVDSTAAPKGSVVVLVP 384
QY 397 -----SQWQFYRLS---GH-----RWASVAAEADRIVGEME 426
DB 395 VGHLEEDRHASQAHLASRNGHISASPPDQPLTPTTEKQDWPAMI SLARKTILSTIQ 444
QY 427 APA-PGFTDSVLDRFIOTPRDIESELGMIGNVMHVMESLDQMKMLRPLPELSHRVPGA 485
DB 445 SRTNVDLTPLIHSTNSPLSKWQTNLDRGAILGLSHSPFNVLCFRPTTRA---RKPGA 501
QY 486 -----DGLVLTGASTHPGGGV 501
DB 502 FDAQLLKFGVLGRAAEVLIIDAFRCGRGDKIKGLVWVGASAPHTGTGV 546

RESULT 7
CRTI_NEUCR
ID CRTI_NEUCR STANDARD; PRT; 595 AA.
AC P21334;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase) (Albino-1
protein).
GN AL-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=90377195; PubMed=21144609;
RA Schmidhauser T.J., Lauter P.R., Russo V.E.A., Vanofsky C.;
RT "Cloning, sequence, and photoregulation of al-1, a carotenoid
biosynthetic gene of Neurospora crassa.";
RL Mol. Cell. Biol. 10:5064-5070(1990).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COPACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- INDUCTION: By photoinduction.
CC -!- SIMILARITY: Belongs to the phytoene desaturase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M57465; AAA33555.1; --
PIR; A35919; A35919;
InterPro; IPR002937; Amino oxidase
InterPro; IPR009150; Bac phytoene dh.
InterPro; IPR002003; NAD_BS.
InterPro; IPR008151; Phyt dehydro.
PFam; PF01593; Amino oxidase; 1.

sphaeroides and overexpression and reactivation of crtII in Escherichia coli and R. sphaeroides." J. Bacteriol. 176:3859-3869(1994).

SEQUENCE FROM N.A.
STRAIN-ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
MEDLINE=95238278; PubMed=7721699;
Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;
"Complete DNA sequence, specific TMS insertion map, and gene assignment of the carotenoid biosynthesis pathway of Rhodobacter sphaeroides"; J. Bacteriol. 177:2064-2073(1995).

SEQUENCE FROM N.A.
STRAIN-ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
MEDLINE=20115911; PubMed=10648776;
Choudhary M., Kaplan S.;
"DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.4.1."; Nucleic Acids Res. 28:862-867(2000).

FUNCTION: This enzyme converts phytoene into zeta-carotene via the intermediary of phytofluene by the symmetrical introduction of two double bonds at the C-11 and C-11' positions of phytoene.

COPACTOR: FAD (Probable).

PATHWAY: Carotenoid biosynthesis.

SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

EMBL; S71770; AAC31138.1; -
EMBL; AJ010302; CAB38739.1; -
EMBL; AF195122; AAF24289.1; -
PIR; S49620; S49620.
PIR; T50745; T50745.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac phytoene_dh.
InterPro; IPR000205; NAD BS.
InterPro; IPR008151; Phyt dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phyt dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
Phylogenesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
Oxidoreductase; FAD; Flavoprotein; NAD.
NP BIND 14 47 FAD (ADP PART) (POTENTIAL).
CONFLICT 73 73 R -> C (IN REF. 3).
CONFLICT 174 174 T -> S (IN REF. 3).
CONFLICT 292 292 L -> P (IN REF. 3).
CONFLICT 395 395 Q -> P (IN REF. 3).
SEQUENCE 518 AA; 57244 MW; 42393515639EF8F1 CRC64;

Query Match 7.4%; Score 203.5; DB 1; Length 518;
Best Local Similarity 22.2%; Pred. No. 1.9e-06;
Matches 135; Conservative 64; Mismatches 201; Indels 209; Gaps 28;

QY 7 AVTVGSGNALVSNALVAREGMSVEVLEKDTVLGGAVSTVERPPGVKVDGSSAHLMIWH 66
Db 13 ALVTGSGGLGAAMRLGAKGWRVTVDKLDVPGRGSSITQ-EGHRFDLGTIVTVV--- 68
QY 67 SGIIIEGLGAHLRLYIDCDPWAPAPPAPOTDGTGIVFHRDLDTATQCSIERACGTKADA 126
Db 69 -----PQSLR---DLWK-----TCG-----RDFADVE-----LKFDP 94
QY 127 YRPFVAVMSERSHVMKASTPTPTGNSNLGAPGLATARGNSLSQFLAPGDALLDVEY 186
Db 95 F--YEVWPGSHFTVRQ-----STEAMKAVAR--LSPGDVAGYKFP 133
QY 187 --DSEALKALAWFGCAQSGPPMSEPGTAPMVGPAALMHVLP-- 226

ProDom; PD139017; Phyt dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
Transmembrane.
NP BIND 11 44 FAD (ADP PART) (POTENTIAL).
TRANSMEM 578 595 POTENTIAL.
SEQUENCE 595 AA; 66367 MW; 0FF3DF07328ED784 CRC64;

Query Match 7.5%; Score 208.5; DB 1; Length 595;
Best Local Similarity 21.7%; Pred. No. 1e-06;
Matches 123; Conservative 79; Mismatches 212; Indels 153; Gaps 27;

Y 7 AVTVGSGNALVSNALVAREGMSVEVLEKDTVLGGAVSTVERPPGVKVDGSSAHLMIWH 66
b 10 ALIVGAGGAGIAPARLAKAGVTVLEKDNFTGRCSLHRTAGTRFDQGPSLLLL--- 66
Y 67 SGI-----IEELG--LGAHLGLYIDCDP-----WAPAPP-----APGTDGPGIVFHRDLDTATCQ 113
b 67 PGLFRETPEDLGTTLEQEDVELLOCFPNYNIW-PSDGKRFSPITD-----NATMK 115
Y 114 -SIERACGTKADAYRPFVAVMSERSHVMKASTPTPTGNSNLGAPGLATARGNSLSR 172
b 116 VEIEKWEG---PDGFRYLSLWLAEGHGYETSL-----RHVLR 151
Y 173 QF-----LAPGDAL-----LDVEYFDSEALKALAWFGAQSQ-PMSEPG 210
b 152 NFKSLLEADPLVVLTLLMALHPESIHWRAGRYFKTDRMQRVFTATMTMGSPFDAPA 211
Y 211 TAPWVGFPAALMHVL--PPGAVGSGGALSA--ALASRMV-----DQATVALGDGVTSTIR 261
b 212 TYSLLQVSELAEGHWYPRG---GFHKVLDALVIGERMVKYRLNTGVSVLTDGKN-G 267
Y 262 RNSNHWTVTTESGREVARKVIACHTILLTDLKNGGPDRTILDHWRKIRVGPICGAV 321
b 268 KKPATGVOLENGVEVLNADLVVNADIVVYTNL-----LPKEIGGI 309
Y 322 LRLATSLPSYRGDATTRESTSGQLL--VSDRAHLTAHGAALAGE----- 366
b 310 KKYANKL-----NNRKASCSISFYWSLSGMAKELETHNI FLAEYKESFDAIPERQA 362
Y 367 LPPRPVILGSPSGIGDPTIAPAGHQVT-----LWSQHPYRLSCHRDWASVAEAD 419
b 363 LPDDPSFYHVPSKVDPSAAPPDRDAVIALVPVGHLLQNGQP-----ELDWPVLVSKARA 417
Y 420 RIVGMEA-----PAPGFTDSVLDRFITQTERDIESELGMITGGNVHVMVMSLDQMLWRPL 474
b 418 GVLTATQARTGLSLSPITTEEI-----VNTFYTWTETKLSKGAILGLAHDFFNVLAFRR 473
Y 475 PELSGHRVPGADGILYLGASTHPPGGV 501
b 474 T-----RAQGMNDNAYFVGASTHPTGTG 495

RESULT 8
RTI_RHOSH STANDARD; PRT; 518 AA.
C P54980; O9RF00;
D 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CRII.
S Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
X Rhodobacteraceae; Rhodobacter.
C NCBI_TaxID=1063;
P [1]
N SEQUENCE FROM N.A.
C STRAIN-ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
X MEDLINE=94292403; PubMed=8021167;
T Lang H.P., Cogdell R.J., Gardner A.T., Hunter C.N.;
"Early steps in carotenoid biosynthesis: sequences and transcriptional analysis of the crtI and crtB genes of Rhodobacter sphaeroides"; J. Bacteriol. 176:3859-3869(1994).

```

RESULT 9
D CRTI_PANAN STANDARD; PRT; 492 AA.
I P21685;
I 01-MAY-1991 (Rel. 18, Created)
I 01-MAY-1991 (Rel. 18, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CRTI.
N Pantoea ananas (Erwinia uredovora).
S Pantoea bacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID:553;
N [1]
N SEQUENCE FROM N.A.
P STRAIN=20D3;
C MEDLINE=91072214; PubMed=2254247;
X Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,
A Nakamura K., Harashina K.;
T "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway
T by functional analysis of gene products expressed in Escherichia
T coli.";
L J. Bacteriol. 172:6704-6712(1990).
C -I- FUNCTION: This enzyme converts phytoene into lycopene via the
C intermediaries of phytofluene, zeta-carotene and neurosporene by
C the introduction of four double bonds.
C -I- COFACTOR: PAD (Probable).
C -I- SIMILARITY: Carotenoid biosynthesis.
C -I- SIMILARITY: Belongs to the phytoene dehydrogenase family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
X EMBL; D90087; BAAL4127.1; -
X F01; D37802; D37802.
X InterPro: IPR002937; Amino Oxidase.

```

carotenoid synthesis by blue light.;

EMBO J. 12:1265-1275(1993).

!- FUNCTION: This enzyme converts phytoene into lycopene via the intermediates of phytofluene, zeta-carotene and neurosporene by the introduction of four double bonds (By similarity).

!- COFACTOR: FAD (Probable).

!- PATHWAY: Carotenoid biosynthesis.

!- INDUCTION: By blue light.

!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M94727; AAA25390.1; -.

PIR; S35306; S35306.

InterPro; IPR000759; Adnck_reductase.

InterPro; IPR002937; Amino oxidase.

InterPro; IPR008150; Bac_phytoene_dh.

InterPro; IPR000205; NAD_Bs.

InterPro; IPR008151; Phyt_dhydro.

Pham; PF01593; Amino oxidase; 1.

PRINTS; PR00419; ADKRDFTASE.

PRODOM; PD139017; Phyt_dhydro; 1.

PROSITE; PS00982; PHYTOENE DH; 1.

Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.

NP BIND 12 45 FAD (ADP PART) (POTENTIAL).

SEQUENCE 529 AA; 58420 MW; 53536A8DFD0D24BC CRC64;

Query Match 6.8%; Score 189.5; DB 1; Length 529;

Best Local Similarity 23.7%; Pred. No. 1.7e-05;

Matches 132; Conservative 74; Mismatches 256; Indels 95; Gaps 25;

8 VVCGSHNALVSAYLAREGWSVEVLEKDTVLGAVS--TVREPCYKVDGSSAHLMI 65

12 IVGAGGGLSALNLAGQGFRTVVSKAVPGRMKGLTLGASGYAVDTGPS---IIQ 68

66 HSGIIEEL-GLGAGHLR-YIDCDPFAFPAPGDTGPGIVFHRDLDTATCQSIERAGTKD 123

69 LPGVLEQIFRRAARLEDDYVKLLPLDWNTRVHFWDGTHLDTTHLDMEAEALAK-FGPRQ 127

124 ADAYRRFVAVWSERSRVHMAKAPSTPPGCSNLIAGFGLATA-----RGNSELSROFLAPG 178

128 ASALRQWEDGEEKYGIAYQKFC--TSADNLGYAPWRAPLRLRPKPMQTLYRQ----- 180

179 DALDEYFDSEALKAAALAWFGAQSPPMSBPGTAPMVGFAALMHVLP-----PGRVAG 231

181 --LDGFFHDDRVTYALAY-----PSKYLGLHPTT-CSSVPSVIPFLAFLGVVHVEG 229

232 GSGALSALASRAVD-GATVALGDGVTISIRNSNHWTTTSGRSHVARKVIAGCHIIT 290

230 GFRELSRGM-MRCARDIGATFRMGTPEVKR-----VDAGRAVGVK--LVGGEVLD 277

291 TLDLGGGDFDRTLDHRRKIRVPGIGAVLRATLSALPSVRGDTATRESTSGIQLLVS 350

278 ADVVVNADLAYAARSILPAEAREGS-----RLTDAALERAKYSCSTFWAYGLDTVYA 331

351 DRAH-----LRTAAGALAG-----ELPP-----RPAVLGMSFGSDIPTAPAGRHV 393

332 DLPHLLIYLSERARTDRDALEDHRVLEDPPFYVNCNPGV-----TDFSGAPAGES-- 382

394 TLNSQMPYRLSGHRDASVAEADRIKVEGMAFA-PGFTDSVLDRFIQTPRDIIESELG 452

383 TLVVLVPTPTGRVDVVKTEQALREKIPAMLEKVLGKGVREHRRERYFTAEWDDFN 442

453 MIGGNVHVEVMSLDQXMLWRPELSGHRVPGAD--GLYLTGASTHPGGGVSG--ASGRS 508

443 VFRGAVENLSHWTQLGLPLRP-----KVKNRDIEGLYFVGGGTHPGSGILTIMESANI 495

509 AARIALSDSRGRKASOW 525

496 AADYLTRAGKGLPGW 512

RESULT 11

ID_CRTI_ERWHE STANDARD; PRT; 492 AA.

AC P22871;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).

GN CRTI.

OS Erwinia herbicola.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.

OX NCBI_TaxID=549;

[1]

SEQUENCE FROM N.A.

RC STRAIN=EH010;

RX MEDLINE=91088634; PubMed=2263648;

RA Armstrong G.A., Alberti M., Hearst J.E.;

RT "Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";

RT Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).

RL !- FUNCTION: This enzyme converts phytoene into lycopene via the intermediates of phytofluene, zeta-carotene and neurosporene by the introduction of four double bonds.

CC !- COFACTOR: FAD (Probable).

CC !- PATHWAY: Carotenoid biosynthesis.

CC !- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M38423; AAA24820.1; -.

PIR; A39273; A33120.

InterPro; IPR002937; Amino oxidase.

InterPro; IPR008150; Bac_phytoene_dh.

InterPro; IPR008151; Phyt_dhydro.

Pham; PF01593; Amino oxidase; 1.

PRODOM; PD139017; Phyt_dhydro; 1.

PROSITE; PS00982; PHYTOENE DH; 1.

Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.

NP BIND 5 38 FAD (ADP PART) (POTENTIAL).

SEQUENCE 492 AA; 54503 MW; 8EDC5DB1562083F2 CRC64;

Query Match 6.8%; Score 179.5; DB 1; Length 492;

Best Local Similarity 22.8%; Pred. No. 7.2e-05;

Matches 129; Conservative 62; Mismatches 239; Indels 135; Gaps 24;

8 VVCGSHNALVSAYLAREGWSVEVLEKDTVLGAVS-TVREPCYKVDGSSAHLMI 67

5 VVIGAGGGLALAIRQAAGIPTVLEQRDKPGR-AYVWHQGTFTDAGPT---VITDP 60

68 GIIEELGLGAGHLRYIDCDPFAFPAPGDTGPGIVFHR-----DLDTCSIERAC- 119

61 TALEALFTLA-GRMEDY--VRLLPVKP-----FYRLCWESGKTLDYANDSALEAQ 109

120 ----GTCDADAYRRFVAVWSERSRVHMAKAPSTPTGCSNLIAGFGLATARGNSEL 175

110 ITQPNRVDVGRFRFLAY-----SQAVFQ-----EGYRLGSPVPLS-----FRDML 151

176 APGDAL-----DEYFDSALKAALAWFGAQSPPMSBPGTAPMVGFAALMHV 223

152 RAGQQLKLQAWQSVQSVSRFTEDHLRQAFSFHLLVG---GNPFTTSSI---YTLIHA 206

224 LPPGRAV---GGGSAALASRAVVDGATVALDGVTSIRNSNHWI-VTTESGREVH 278
 207 LERENGWFFPEGCTGALNGVMKLEFDLGGGIELNARVEELVWADNRVSQVRLADGRIFD 266
 279 AKVITAGCHILTLDLGNGGDFRITLDHMRKIRVGPICGAVLRLATSALPSVRCDAIT 338
 267 TVAVASNDVNTYKL-----LGHHPVGOKRAALE 298
 339 RESTSGLOLV---SDRAHLRTAH-----GAALAGELF---PREAVL 374
 299 RKSMSNSLVLYFLGNQPHSLAHTTCFGRYRELIDEIFTGSAADDSFLVLSHSPV- 357
 375 GMSFGIDPTTAPAGCHVTLWSQVRLSGHSDWASVAREADRIVGEM-EAPADGFT 433
 358 -----TQPSLAPPCASFYVLA-PVHLGNAPLDWAQEGPKLDRIDFYLEERYMPLGR 410
 434 DSVLDRFIQTPRDIESELGMGNVHMVMSLDQMLWRPULPELSGHRVPGADQLYLTA 493
 411 SOLVTQRIPTADFDHTDLAHLGSAFSLEPLLTQSANFRP-----HNRDSIANLYLVA 465
 494 STHPGGVSG--ASGRSAARTALSD 516
 466 GTHPGAGIPGVVASAKATASLMIED 490

SULT 12

TD_RHOCA
 CRTD_RHOCA STANDARD; PRT; 494 AA.
 P17059;
 01-AUG-1990 (Rel. 15, Created)
 01-AUG-1990 (Rel. 15, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Methoxyneurosporene dehydrogenase (EC 1.1.4.99.-).
 CRTD.
 Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 Rhodobacteraceae; Rhodobacter.
 NCBI_TaxID=1061;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=SB1003 / St. Louis, and BEC404;
 MEDLINE=89313663; PubMed=2747617;
 Armstrong G.A., Alberti M., Leach P., Hearst J.E.;
 "Nucleotide sequence, organization, and nature of the protein
 products of the carotenoid biosynthesis gene cluster of Rhodobacter
 capsulatus.";
 Mol. Gen. Genet. 216:254-268 (1987).
 [2]
 SIMILARITY TO CAROTENOID DESATURASES.
 MEDLINE=90368827; PubMed=2144293;
 Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
 "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
 crassa are structurally and functionally conserved and contain
 domains homologous to flavoprotein disulfide oxidoreductases.";
 J. Biol. Chem. 265:16020-16024 (1990).
 -!- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR
 METHOXYNEUROSPORENE TO SPHEROIDENE.
 -!- COFACTOR: FAD (Probable).
 -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 -!- SIMILARITY: Belongs to the phytyene dehydrogenase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; X52291; CAA36537.1; -;
 EMBL; Z11165; CAA7544.1; -;
 PIR; S04406; S04406.

DR InterPro: IPR002937; Amino oxidase.
 DR InterPro: IPR008150; Bac phytoene dh.
 DR InterPro: IPR008151; Phytin dehydro.
 DR Pfam: PF01593; Amino oxidase; 1.
 DR ProDom: PD139017; Phytin dehydro; 1.
 DR PROSITE: PS00982; PHYTOENE_DH; 1.
 KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 KW Oxidoreductase; Flavoprotein; NAD.
 FT NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 494 AA; 52312 MW; D1180A023FFEB5A9 CRC64;
 Query Match 6.4%; Score 176; DB 1; Length 494;
 Best Local Similarity 23.4%; Pred. NO. 0.00012;
 Matches 132; Conservative 55; Mismatches 254; Indels 122; Gaps 26;
 QY 1 KSAFLDADVVGSGHVALVSAAAYLAREGWSVEVLKNTVLGCAVSTVERFPGYKYVDRGSSA 60
 DB 1 MRSETDVVVGARMGSLAAAGAAAGLVTIVVEAGDAPGKARAVPT-PGGPADTGTPTV 59
 QY 61 HLMIRHSIIIEEL--GLGAHLRYIDCDP-----WAFAPPAGTGGPGIVFHRDLDA 110
 DB 60 -LTMFH--LDALFAACGTAAEHLTLIPLRLAHFM-----PDGSLDLFTDTTEA 108
 QY 111 TQSIERACGTDADAVRRF---VAVMSRSRHWKAFSTPTGSLNIGAFGLATARG 166
 DB 109 NIEAIRAFAGDKAAAFRRFDHLTTGLWEAFHRSVIAA---PKDLMRIIA----- 156
 QY 167 NSELSRQFLAPG-----DALDEYFDSALKKAALAWFGAQSOPMSEPGTAPMVGFPA 219
 DB 157 -ATVTRPOLWPAIRPGLTMRDLAHFFKDPRLAQLFCRYATYVG---GRPGATP-----A 207
 QY 220 LMHVLPPCRAGV-----GSGALSAALASRAVVDGATVALDGVTSIRNSNHWI-VTTE 272
 DB 208 VLSLIWQAEVQGWALREGMHVAALAAARVAAKGVRFHYGAKAKIRVKEGRTAVEIE 267
 QY 273 SGREVHARKVIAGCHILTLTD-LGNGGDFRITLDHMRKIRVGPICGAVLRLATSALPS 331
 DB 268 TGVSIPTGCACIFNGDPCALRDGLIGDAA--RASME---KSPRPAPSLSAWV-WAFGATP- 320
 QY 332 YRGDATTRESTSGLOLV---VSDRAHLRTAHGAALAGELPRPAVLGMSFGIDPTTAPA 388
 DB 321 -----IGVDLAHNVFVTTADPELFGPIGAGEMEPEETLY-----ICAQ 359
 QY 389 GRHQVTLWSQVRYL-----SGHRDWSVAEAEADRIVGEMAEAFAPGFTDSVLDRIOT 443
 DB 360 DREMQAPVPEIEREELIMNGPAGHPQPPQEEAQCARTFPMLAAMGLTFSPDPETRALTT 419
 QY 444 P-----RDIESELGMGNVHMVMSLDQMLWRPULPELSGHRYP---GADGLYLTGAST 495
 DB 420 PALLSRFPQSLGAIYGG-----SPEGLTATFRPLARTGLKGLYLAGGTT 465
 QY 496 HPGGGVSGA--SGRSAARIALSD 516
 DB 466 HPGAGVPMALTSQTHAARALLAD 488
 RESULT 13
 ID CRTI_PHYBL STANDARD; PRT; 583 AA.
 AC P54982;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.1.4.99.-) (Phytoene desaturase).
 GN CARB.
 OS Phymyces blakesleeanus.
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
 OC Phymyces.
 OX NCBI_TaxID=4837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 1555;
 RA Ruiz-Hidalgo M.J.;

```

Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the intermediary of phytylfuene by the symmetrical introduction of two double bonds at the C-11 and C-11' positions of phytoene.
-!- COPFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; X78434; CAAS5197.1; -.
PIR; S43139; S43139.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR00205; NAD_BS.
InterPro; IPR008151; Phyt_n_dehydro.
Pfam; PF01593; Amino oxidase; 1.
ProDom; PD139017; Phyt_n_dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD; Transmembrane.
NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
TRANSMEM 531 551 POTENTIAL.
SEQUENCE 583 AA; 65983 MW; B0B8F682B12FB591 CRC64;

Query Match      6.4%; Score 176; DB 1; Length 583;
Best Local Similarity 21.2%; Pred. No. 0.00015;
Matches 120; Conservative 68; Mismatches 219; Indels 158; Gaps 23;

y    8 VVVGSHNALVSAYLARQNSVEVLKDTVGLGAVSTVERPPGYKDYRGSSAHLMLRH 67
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
b    8 ILICAGAGGTATAARLAREGIKVTVYKNFGGRCSLINH-NGRFDQPSLYLPK-- 64
                                         |::|::|::|::|::|::|::|
y    68 GIIEELG-----LGAHGLRVLDCCPWAFAPPACTDGEIVFHFRDLDATQS 114
b    65 -LFEEAFEALDEKHEDVELLRCHNYKVHF-----DDXKLQLSSDSLNRKPE 112
                                         |::|::|::|::|::|::|::|
y   115 IERACGTDADAYRRFVAVWSERSHYWKAFSTPTPGSNLICAFGGLATA-RGNSE--- 169
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
b   113 MERIEG---PDGFLRFDEMGSHTHYE-----GGVEMLAKQNFEIWK 153
                                         |::|::|::|::|::|::|::|
y   170 LSRQFLAPGDALLD-----EVFDSEALKAAIANFGRQSG-PNNEPGT----- 211
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
b   154 LISLVYPALFRLHIHDFVYSRAAKYFTKKRMMAFTQSMYMGSPVDSPAVNLLOYT 213
                                         |::|::|::|::|::|::|::|
y   212 -----APMVGFAALMHVLPFGRAVGGSGALSAAASRMVADGATVAL-----GDGVTS 259
b   214 EPAEGIWPKGFNTVIQKL-----ENIATEKFGARFIIVEAPVAXINTDDKGVKG 265
                                         |::|::|::|::|::|::|::|
y   260 IRNSNMHWVTTESGREVHARKVIAGCHIL-----TTDLLNGGFDRITLD-H 307
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
b   266 -----VTLQSGRVIADAWVCNADIIVAYHNLLPPCRWNTTNLEKKLTSSISFY 316
                                         |::|::|::|::|::|::|::|
y   308 WRKIRVGPIGA-VLURLATSALPYRGDAFTRESTSGLQLLVSDRAHLRTAHGAALAGE 366
b   317 MSLK-RVPFELDYNHTFLAEAKESPFDELFTDKHPSELSPVYN----- 359
                                         |::|::|::|::|::|::|::|
y   367 LPRPRAVLGMFSFGIDPTTAAGR-----HOVTLSQWQPYLRSRHWDVASVAEA 416
b   360 LPSR-----IDFTAAPFKDSMIVLVPIGHMKSKITNEADYTNI VKRAKRYLVE- 408
                                         |::|::|::|::|::|::|::|
y   417 EADRIVCEMEAFAPGGTSDVLDRFIQTPTREDISELGMIGGNVMHVMSLDQMWRPLPE 476
b   409 -----VLERRLGUTNFIDLVEHEVNDPIWKKFNLMWRGSLIGLSDYLVLFRRSTQ 463
                                         |::|::|::|::|::|::|::|
y   477 LSGHRVPGADGLYLTCASTHPPGGV 501

```

464 DSTGRY---KNLFPVGCASHTPGTGV 485

RESULT 14
CRTI AGRAU STANDARD; PRT; 501 AA.

ID AC P54978:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhinobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=414155;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=96062243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,
RA Ohtani T., Miki W.;
RT "Structure and functional analysis of a marine bacterial carotenoid
biosynthesis gene cluster and astaxanthin biosynthetic pathway
proposed at the gene level.";
RL J. Bacteriol. 177:6575-6584(1995).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
intermediates of phytyluene, zeta-carotene and neurosporene by
the introduction of four double bonds (By similarity).
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
biosynthetic pathway.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: D58420; BAA09594.1; .
InterPro: IPR002937; Amino oxidase.
InterPro: IPR008150; Bac.phytoene dh.
InterPro: IPR008151; Phyt.dnhydrol.
Pfam: PF01593; Amino oxidase; 1.
ProDom: PD139017; phyt.dnhydrol; 1.
PROSITE: PS00982; PHYTOENE DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
XP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SEQUENCE 501 AA; 54806 MW; 5F251AF11D679358 CRC64;

Query Match 6.3%; Score 174.5; DB 1; Length 501;
Best Local Similarity 24.1%; Pred. No. 0.00016;
Matches 135; Conservative 51; Mismatches 251; Indels 123; Gaps 24;

QY 7 AVVVGSGHNLVSAYLARCGNSVEVLKDTVLGGAVSTVERFGYKYVDRCSSAHLMIKH 66
DB 11 ALVIGAGGGLALAIRQSAGIATTVLEARDKPGR-AVVWHDQGHLPAGEPT---VITD 66
QY 67 SGITISELGLGAHLYRITDCDFWAPAPPFTGD---GPQIVFER-----DLNACQS 114
DB 67 PDALKEL-----WALTQDNARDVTLMPSVPFYFLRWPGPKVFYVNEADQ 112
QY 115 TERAC---GTQDAAYRRFVAVWSERSRHVMKAFSTPPGTGSNLICAFGLATARGNSELS 171
DB 113 LERQIAQFNPDLEGYRFDRDYAEVYQSGYVKLGTVP-----FLUKCQMKA 160
QY 172 RQLAPGDDALIDEYFDSEALKALAWFFQAQ-----SGPMSBFGTPMVGPF 217
DB 161 ----APALMKLEAY--KSVHAKVAFTIKDPYLROAFSVHTLLVNGNFSTS-----SI 207
QY 218 AALMHVLPQGRAV-----GGSCALSALLASRMADVGTALVGCVTSIRRNHNWT-VTTE 272

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

! protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 14.808 Seconds
(without alignments)
3455.835 Million cell updates/sec

tle: US-09-941-947a-38

irect score: 2768

quence: 1 MSAPLDVVVSGHNALVSA.....ALSDSRGKASQWRRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tal number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	859.5	31.1	511	2 E75561	probable phytoene
2	688	24.9	542	2 S76617	hypothetical prote
3	660.5	23.9	565	2 A12273	hypothetical prote
4	538.5	19.5	543	2 G87635	phytoene dehydroge
5	533	19.3	518	2 G20413	phytoene dehydroge
6	466.5	16.9	544	2 T32568	hypothetical prote
7	439.5	15.9	538	2 B72609	hypothetical prote
8	418	15.1	489	2 F75591	probable dehydroge
9	340	12.3	472	2 T35936	probable dehydroge
10	327	11.8	469	2 S09189	49K protein - Stre
11	314.5	11.4	536	2 A70523	probable oxidoredu
12	303	10.9	535	2 F70782	probable oxidoredu
13	284	10.3	473	2 C70915	hypothetical prote
14	283	10.2	480	2 A70854	hypothetical prote
15	275	9.9	523	2 T36968	probable phytoene
16	270	9.8	501	2 T37591	hypothetical prote
17	264.5	9.6	517	2 T31463	probable diaphophyt
18	249	9.0	512	2 H84320	phytoene dehydroge
19	247	8.9	536	2 B84327	phytoene dehydroge
20	229.5	8.3	528	2 T51119	phytoene desaturas
21	225	8.1	582	2 T46822	phytoene desaturas
22	224	8.1	495	2 S23633	methoxyneurosporen
23	224	8.1	621	2 T48646	phytoene dehydroge
24	222	8.0	485	2 S49624	methoxyneurosporen
25	214.5	7.7	470	2 T72567	hypothetical prote
26	212.5	7.7	506	2 T34912	probable carotenoi
27	212	7.7	430	2 F90272	phytoene dehydroge
28	211	7.6	380	2 A12185	hypothetical prote
29	208.5	7.5	595	2 A35919	carotenoid biosynt

30 204.5 7.4 518 2 T50745 phytoene dehydroge
31 203.5 7.4 518 2 S49620 phytoene dehydroge
32 202 7.3 485 2 S32171 hydroxyneurosporen
33 200 7.2 511 2 T50910 phytoene dehydroge
34 197.5 7.1 548 2 C75466 phytoene dehydroge
35 191.5 6.9 492 2 D37802 phytoene dehydroge
36 191 6.9 486 2 T50749 methoxyneurosporen
37 189.5 6.8 529 2 S35306 phytoene dehydroge
38 189 6.8 506 2 A82064 hypothetical prote
39 188 6.8 430 2 E84212 hypothetical prote
40 187.5 6.8 492 2 S52586 phytoene dehydroge
41 181 6.5 587 2 A86203 hypothetical prote
42 179.5 6.5 492 2 A33120 phytoene dehydroge
43 176.5 6.4 503 2 A24446 hypothetical prote
44 176 6.4 494 2 S04406 methoxyneurosporen
45 176 6.4 583 2 S43139 phytoene dehydroge

ALIGNMENTS

RESULT 1

E75561

probable phytoene dehydrogenase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: E75561

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Na
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: E75561

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-511 <WHI>

A:Cross-references: GB:AE001872; GB:AE000513; NID:g6457750; PIDN:AAF09686.1; PID:g645775

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0093

A:Map position: 1

C:Superfamily: phytoene dehydrogenase

Query Match	31.1%	Score 859.5;	DB 2;	Length 511;
Best Local Similarity	39.7%	Pred. No. 3.1e-51;		
Matches 207;	Conservative 80;	Mismatches 208;	Indels 27;	Gaps 11;
QY	6	DAVVVSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERFPFGYKVDGSSAHLMI	65	
DB	5	DLIVMGAGHNALVTAAYAAARAGLVGVFERRHLVGGAVSTEEVVPVGYRFDYGGSAHLIL	64	
QY	66	HSGTIEELGLCAHGLRVIDCDPMAPAPPAGTDPGPIVF-HRDLATCQSI-ERACGTXD	123	
DB	65	MTPVIRELELTHGLHVLVDPMFHA-----SDGETPFIHRDAGRTIRELDEKFPG-Q	117	
QY	124	ADAVRFVAVWSESRHMKAPSTPTPGSNLIGAFGLATARG-----NSELRFQFLAPG	178	
DB	118	GDVYGRFLDWTTPPARAVADLFNSAPGLDL-----GKQVNRSGQGDWNSQLPR-ILRPY	172	
QY	179	DALLDEYFDSKAAALAWCAQSGPMSPEPTAPMVGFAALMHLVPPGAVGSGSALSA	238	
DB	173	GDVAREFSESRVAPATITWAAQSGPPSPFLSAPFLLMHPLYHEGVARPKGSGGLTK	232	
QY	239	ALASRMAVDGATVAGDGVTSIR-RNSNHMTVTTEGSRVHAKVITAGCHLITLILGN	297	
DB	233	ALRRATEAGGGEVFTDAPVKVILVKDGAQIRLESGETVTARAVSGVHILITANAL--	290	
QY	298	GGFDRITLIDHWRKIRVGPGLVRLATLSALPSYGDATTRETSGLQLLVSDRAHLRT	357	
DB	291	---PAEYVPSAARNVRVNGFGMLLRLLALSEKVKYIR-HTEPDSRIGLGLIINERQIMQ	346	
QY	358	AHGAALAGELFPPEAVLGMSPSGIDPTIAPAGRHQVTLWSQMPYRLSGHRDWSVAEAE	417	

347 GYGEYLAGQETDPLVAMSPFSAVDSDSLAPPNGDVLWLAQYYPPELA-TGSWEITAE 405
418 ADRTVGMERAPQFTOSVLDRTQTQTRDIESELGMLGNNVMEVMSLDQMLRLPLPEL 477
406 RENTLAFHYAPGTRTIVGELVQTEPQWLETNLGRGNNVMEVMSFDQMFSPFWLKA 465
478 SHRVPGADGLYLTGASTHGGVSGASRAARIALSDSRR 519
466 SQVRWPGVQGLYLTGASTHGGVSGASRAARIVKDLTR 507
RESULT 2
176617
hypothetical protein - Synecocystis sp. (strain PCC 6803)
Species: Synecocystis sp.
Variety: PCC 6803
Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
Accession: S76617
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
NA Res. 3, 109-136, 1996
Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
Reference number: S74322; MUID: 97061201; PMID: 8905231
Accession: S76617
Status: preliminary
Molecule type: DNA
Residues: 1-542 <KAN>
Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BAAL0561.1; PID:g100172
Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
Superfamily: phytoene dehydrogenase
Query Match 24.9%; Score 688; DB 2; Length 542;
Best Local Similarity 31.9%; Pred. No. 1.8e-39;
Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;
6 DAVVSGSNALYSAAVLAREGMSVLEKDTVLGAVST----VERPPGYKVDGSSAH 61
5 DVVIGAGHGLVCAAYLLQRLGVLTLKREVPVGGATTEALMPELSPFPNRCALDH 64
62 LMRHSGITBELGAGHLYRIDCDPAWAPAPGTDGPGIVFHRDLDTQOSIERACGT 121
65 EFTLPGVLELQAQGLYFLCDPSVF---CPGLDQAPMSYSLKTKAHI-ATYSP 120
122 KDADAYRRFVAVWSERSRHVKAFSTPTG-----SNLIGAPGLATARGNS 171
121 RDAEKYRQFVNYWIDLVNAVQPAFNAPPQALLDLALNYGWNKSLVLAIGSKTKALDFI 180
172 RQFLAPCDALLDFYDSEALKALAWFGAQSGPPMSEPCGAPMVGFALMHVLPFGRAVG 231
181 RTMIGSPEDVLNWFDSERVKAPLARCSEIGAPPSQKSSGSSGMMVAMRHLEGIARPKG 240
232 GSGALSAALASMAVNDGATVALGDGVTIR-RNSNHWTTTSGREVHARKVIACCHILT 290
241 GTGALTALVQLVQAQGGKILDTQTKRVLVNNOAIGVEVANGEQYRAK-----GVIS 295
291 TLD-----LLGNGGDFRTTLDHWR-KIRVPGIGAVLR--ATSALPSYRGDATTRE 340
296 NIDARLFLQLVEPGALAKVQNGLSERLRTVNNNEALKIDKALSGLPHTAVAGP-E 354
341 STSGLOLLVSDRAHLRTAAGALAGELPRPVLGMSF-SGIDPTIAPAGRHQVTLMSQ- 398
355 DLGTGTLIADSVRHHVBEAALIALGOIPDANPSLYLDIPTVLDPMTAPPGQH--TLWIEF 412
399 WQPYRLSG-----HRDWA-SVAEAEADRVIGMEAFAPGFTDSVLDRTFOTPRETSE 450
413 PAPYRIAGLEGTGLMGTGWTDELKVKVADRVLDKLTDPAPNKLKSLIIGRVSPAEIAQR 472
451 LGMIGNVHVMVMSLDQMLWPLPELSGRHVFGADGLYLTGASTHGGVSGASRAA 510
473 LGSYNGVNYHLDMSLDQMLPLPELPIANYQTP-INKNYLTGAGTHPGGSGISGMPGNCA 531

511 RIALSDSRR 519
532 RVFLKQQR 540
RESULT 3
AI2273
hypothetical protein all3744 [imported] - Nostoc sp. (strain PCC 7120)
Species: Nostoc sp. PCC 7120
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AI2273
R; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An.
A:Reference number: AB1807; MUID: 21595285; PMID: 11759840
A:Accession: AI2273
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-565 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW75443.1; PID:g17132878; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all3744
C:Superfamily: phytoene dehydrogenase
Query Match 23.9%; Score 660.5; DB 2; Length 565;
Best Local Similarity 31.7%; Pred. No. 1.4e-37;
Matches 174; Conservative 94; Mismatches 224; Indels 57; Gaps 17;
6 DAVVSGSNALYSAAVLAREGMSVLEKDTVLGAVSTV2----RPPGYKVDGSSAH 61
5 DVVIGAGHGLVCAAYLLKAGYSVLLKRSVPGGAATTECLPKPEAGFPKFNCAIDH 64
62 LMRHSGITBELGAGHLYRIDCDPAWAPAPGTDGPGIVFHRDLDTQOSIERACGT 121
65 EFTLPGVLELKGKYLECDDPVE---CPHPDGKYFLAHSLEKTCAEIAR-YSE 120
122 KDADAYRRFVAVWSERSRHVKAFSTPTP-----TGSNLIGAPGLATARGNS----EL 171
121 RDAKYAEFTYQRAIGAMIPFNAPPKSIIDVGNVDITKPLFSGVSGPNKTLDFI 180
172 RQFLAPCDALLDFYDSEALKALAWFGAQSGPPMSEPCGAPMVGFALMHVLPFGRAVG 231
181 RNMLTSABDIINWFSEFELKAPLARLASELGAAPPQKTIAGAIMAMRHNPCHAPRG 240
232 GSGALSAALASMAVNDGATVALGDGVTISRRNSNHWTTTSGREVHARKVIAGCH---- 287
241 GTGALIKALVNLVSKGVILTDQVEK-----VLIDDGKAVGR-VSGGTEYRAK 290
288 --TLTLD-----LLGNGGDFRTTLDHWR-KIRVPGIGAVLR--LATSALPSYRGD 335
291 YGVISNIDAKRFLQMTDKSDVDAADPDQWELERIRVNNNETILKIDLDLDEPLRFPFH 350
336 ATTRESTGLQLLVSDRAHLRTAAGALAGELPPR-PAVLGMSFGSDIPTIAPAGRHOVT 394
351 AHRDEYLVGSLIADSVHAEQAKSKTIGEIPSDSPSMYVMPSYLDPTLAPSKH--T 408
395 LMSQ-WQPYRLSGH-----DMSVAEAE-ADRVIGMEAFAPGFTDSVLDRTFOTPR 445
409 WMIEFPAYQIAGAGTGKGTGWTDELKNQVADKLVQKLYATYAPNVKTATIAREVSPA 468
446 DIESELGMITGGNNVMEVMSLDQMLWPLPELSGRHVFGADGLYLTGASTHGGVSGAS 505
469 ELGERLKGKNNYHLDQMLWPLPELPIANYKTP-IDNLYLTGAGTHPGGSGISQMP 527
506 GRSARIAL 514
528 GRNCARVFL 536
RESULT 4

37635

ytocene dehydrogenase-related protein [imported] - Caulobacter crescentus

Species: Caulobacter crescentus

Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

Accession: G87635

Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

3.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolod

J.; Brumby, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

oc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

Title: Complete Genome Sequence of Caulobacter crescentus.

Reference number: A87249; MUID:21173698; PMID:11259647

Accession: G87635

Status: preliminary

Molecule type: DNA

Residues: 1-543 <STO>

Cross-references: GB:AE005673; NID:gl3424781; PIDN:AAK25083.1; GSPDB:GN00148

Genetics:

Gene: CC3121

Query Match

Best Local Similarity 19.5%; Score 538.5; DB 2; Length 543;

Matches 180; Conservative 76; Mismatches 234; Indels 53; Gaps 21;

6 DAVVVGSHNALSAAVLAREGWSVEVLEKDTVLGGAVSTVERPGYKVDGSGSAHLMIR 65

8 DAVIIGGSHNGLVCAFLAKAGLVTVCEARGVVGGAATVEEFHPGPR-NSVASYTVSLL 66

66 HSGIIEELGLGAGHGLRYIDCDPWAFAPAPGTPGPIVFRHDLDTATCQSIERACGTAKD 125

67 NRVFIADMLGRLGLTFLERISNFIPI--SDDKIYKLGGLERT-QEERFYKSRDAE 122

126 AYRRFVAVWSRSHVMK--AFSTPTT-GSNLIGAPGLATARGNSLSRQ-----P 174

123 VLPAYYAMLDEIG-DILDLAQSTPPNLDGLPGLLRALRQGGRLAFLSRKRDLLDLF 181

175 LAPGDALDDEYFDSALKAALAWFGAQSGPPMSEPTAPMVGFPAALMHLVLP----- 256

182 TKSARDVLGWFSEFPVKAAG-EDAVVG-NFASPTD-PSAYVLLHETGEVNGKGAW 238

227 GRAVGGGALSAALASRAVMDGATVALGDGVTSTIRNSNHW-VTTSGREVHARKVIAG 285

239 CHAVGGMGAITQAKAKACEAAGVEILLDAPVEAVHIDGGKAAGVQLVDGRQIMAPIVAN 298

286 CH-ILTLDLGLNGGFD---RTLDHWRKIRVGPVGAV-LRLATSALPSYR---GDAT 337

299 VNPALLYKKLVPPSALTDFPKAVDGYKN-----GSGTFRMVVALSEFTCLPKET 352

338 TRETSGLQLIVS---DRAHLRTAAGALAGELPPRPAVLGMSFGSIDPTIAPAGEHQV 393

353 AEHQSGIIVAPSLDYMDAAV-RDAKGGIS---KAPIVEMLIPLSSLDTSLAPPGQHVA 407

394 TLWQWQPYRLSGHEDNASVAEADRIVGEAEAFAPGFTDSVLDRIQTPTDIESELGM 453

408 SLFCQQFAPFLPDGRGSDWADEAAADLIIDTVQWAFPKASVGLGRMILSPDLERKFLG 467

454 IGGNVHVMESLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSGASGRSAARIA 513

468 IGGDLMHGHMSLDQLWATRLPLGHASHRAPTA-GLYWCAGTHFGGVSNGPGRNAREI 526

514 LSD 516

527 LRD 529

SULT 5

0413

ytocene dehydrogenase related protein [imported] - Sulfolobus solfataricus

Species: Sulfolobus solfataricus

Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

Accession: G90413

She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awevez, M.J.; Chan-

ong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P

rett, R.A.; Ragan, M.A.; Sensesen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90413

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-518 <KUR>

A:Cross-references: GB:AS006641; NID:gl3815726; PIDN:AAK42566.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO2422

Query Match

Best Local Similarity 19.3%; Score 533; DB 2; Length 518;

Matches 168; Conservative 93; Mismatches 226; Indels 56; Gaps 21;

QY 5 LDVVVGSHNALSAAVLAREGWSVEVLEKDTVLGGAVSTVERPGYKVDGSGSAHLMIR 64

Db 2 IDVAIIGGSHNGLVTAAYLAKAGLVAVFERREIVGGASVTEELWPSIKYSTGAYVLSLL 61

QY 65 HSGIIEELGLGAGHGLRYIDCDPWAFAPAPGTPGPIVFRHDLDTATCQSIERACGTAKD 124

Db 62 RPK-IIEELKLRERGLKVYKDPGLF---LPFENGKKLYIWSSEKTKKEIEK-PSKND 116

QY 125 DAYRRFVAVW---SERSRHVMKAFSTPTT---TGSNLIIGAPGLATARGNSLSRQ-----RQF 174

Db 117 KNYKQWVRFMDLFAEMADFFM--LNPPPHIDEASNLINIFRG---NYNEELALSPLTF 170

QY 175 LAPGDALDDEYFDSALKAALA---WFGAQSGPPMSEPTAPMVGFPAALMHLVLP----- 226

Db 171 MQDAKSLDDEYFETDEVKSALEDSVVGTPASP-STFGTA---YVLAHVIGEIVNGVK 224

QY 227 ---GRAVGGGALSAAL---ASRMVADGATVALGDGVTSTIRNSNHWVTTSGREVHARK 280

Db 225 GAWGYVEGGMGEVTKALRSASHLGVSIYTNARDEV--LVNKRVEGIGKLNKGTINAK 282

QY 281 KVIAGCHILTT-LDLLNGCGFDRITLDHWRKIRVGPVGIGAVLRLA--TSALPSYRGDAT 337

Db 283 IIVSNADPTTFKKLRNALEEDFL----RRVALKXGVGSFKIVGLELPLDFGCKS 338

QY 338 TRETSGLQLIVSDRAHLRTAAGALAGELPPRPAVLGMSFGSIDPTIAPAGEHQVTLWS 397

Db 339 LSPHIASELIMPNEYIEKAYDDARALGYSREPLWLSINIQSSVDPTVAPPKPSFSG 398

QY 398 QWQPYRLSGHEDNASVAEADRIVGEAEAFAPGFTDSVLDRIQTPTDIESELGMGGN 457

Db 399 QYLVVD-SKRD--EMEKIAEITFEKIEFAPNFKIYE--VLTPLDIERFRFGINGN 453

QY 458 VNHVMSLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSGASGRSAARIALSDS 517

Db 454 IFHLDMTDQLYFRPLIGYSDYTP-IRGLYLCGSGTHPGGVTGAGYNHVKILEDL 512

QY 518 RRG 520

Db 513 RRG 515

RESULT 6

T32568

hypotheical protein F37C4.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T32568

R:Langston, Y.; Sansone, J.; Wohldmann, P.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid F37C4.

A:Reference number: Z21193

A:Accession: T32568

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-544 <LAN>

A:Cross-references: EMBL:AF036705; PIDN:AA095172.1; GSPDB:GN00022; CESP:F37C4.6

A:Experimental source: strain Bristol N2; Clone F37C4

C:Genetics:


```

y 61 HLMIRHSGIIEELGAGGLRYIDCDPWAPAPAP---GTDGPGIVFHRDLATCSIE 116
b 77 HPLAASAPAPREWLEAFGLRIH-----PPAPLQGTLAGGSGVTLERDLSATA--- 125
/ 117 RACGKTODAYR-----FVAVMSRSRHVKAPSTPTGSMNLIAGFGLATARGNSLSR 172
c 126 -AVIAGDGAWEELPAPLQVEEELDDILRPLPHVPHPTFLAREGLRA----- 174
/ 173 QFLAPCDALLDEYFSEALKALAWFGAQSGPPMSEPGTAPMVGPAAL-MHVLPPGRAVG 231
c 175 --LPPADLGLRTFTPQARALWXGIAHTGLPLTPTGSAITLVALTAVAGWPPFAG 232
/ 232 GSGALSAAASMAVDGATVALGCGVTSIRNSNHWTTTSGREV-HARKVIAGCHILT 290
c 233 GAQALADALRYLYELG-----GEVITGR-----VSQPRDLPPARTVLDSSPAV 278
/ 291 TLDLNGGDFRTTLDHWR--RKIRVPGICGAVLRLATSLP-----SVRGDATTRETS 344
c 279 LRLGLG-----DRAPASYRAALSRYVYAGMCKFYALSGPLPWQDERLRAATVH 334
/ 345 LQLLVSDRAHLRTANGAALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQ 404
c 335 ABIVASE-----ASIQSE---RPYLLAAQHSLEFDSRAPAGGH--TLWV----- 373
/ 405 SGHRDASVAB-AEAD---RIVGMEAPAGFTSDVLDRTIOTPRDIES-ELGMIGNM 459
c 374 -----YAHVNGSDADLRPRVEAQLERFAPGHEKVLACRVITASQLERSPV 428
/ 460 HYEMSLDQMLWRPLPELSGHRVPGADGLYTGASTHPGGVSGASGRSARIALSD 516
c 429 GGAATLPQOVA-RPVLSAAPTYP-VRGVYLCSSSTPPGGGIGHGCGVNAARTALAD 483

RESULT 9
15936
:obable dehydrogenase - Streptomyces coelicolor
:Species: Streptomyces coelicolor
:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-Jan-2000
:Accession: T35936
:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
:Submitted to the EMBL Data Library, January 1999
:Reference number: Z21551
:Accession: T35936
:Status: preliminary; translated from GB/EMBL/DBJ
:Molecule type: DNA
:Residues: 1-472 <SEE>
:Cross-references: EMBL:AL035206; PIDN:CAA22758.1; GSPDB:GN00070; SCODB:SC9B5.16
:Experimental source: strain A3(2)
:Genetics:
:Gene: SCODB:SC9B5.16
:Superfamily: phytoene dehydrogenase

Query Match 12.3%; Score 340; DB 2; Length 472;
Best Local Similarity 27.8%; Pred. No. 9.6e-16;
Matches 149; Conservative 63; Mismatches 226; Indels 98; Gaps 24;

1 MSALFAYVVGSGHNAVSAAYLAREGMSVEVLEKDTVLGGAVSTVE-RFPYKVDGSS 59
1 MFSLMDLVVVGAGPNGLTAVALARRGPPVAFVFAQGTGGGARTTELTLPGRHDP 60
60 AHLIRHSGIIEELGAGGLRYIDCDPWAPAPAGTGDGPGIVFHRDLATCSIERAC 119
61 AHLPLGINSFAPRGLFLERYGLEWHL-PGLPMAHPFP--DGSAAVLSRVGETAASF 113
120 GTKODAYR-----RFVAVMSRSRHVKAFSTPTGSMNLIAGFGLATARGNSLSR 175
114 GADAGPFRVLVERFLPNDTLARDPM---SLPIT-----ALPRDPTVLARFGL 159
176 A--PGDALLDEYFSEALKALAWFGAQSGPPMSEPGTAPMVGPAALMHVLPGR 233
160 VGLPSTWLMRRFRDEKAKTLFAGLVAVH-----MAPLGGFAT-----GAIGLV 203
234 GALSAAASMAVDGATVALGCGVTSIRNSNHWTTTSGREVH-----AR 280

```

```

Db 204 FALAARHGWVPARGSSQSSISDALTYLK-----DLGAVHTDYEVRKLDLDP 254
Qy 281 KVIACGHILTTLDLNGGDFRTTLDHWRKIRVPGICGAVLRLATSLPSVRGDATTRE 340
Db 255 AYVFTSPFALARIAGLG-----SHY-ANRYGPSVKIDYALDGPVP-----WTAE 301
Qy 341 STS-GLQLLVSDRAHLRTA-HGAALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQV 398
Db 302 PRSAGTVQIGADSTGTEIGALHAASOTDRAPERPFLLITVQPSVADFTRAPAGKH--VF 359
Qy 399 WQPYRLSGH--RDNASVABEADRVIGMEAPAGFTSDVLDRTIOTPRDIES-ELGMIG 455
Db 360 -----GHVPGNMGDL---TDAMERQLERFAPGRDRLARATAGPAELAAARNANT 409
Qy 456 GNMVHMSLDQMLWRPLPELSGHRVPGADGLYTGASTHPGGVSGASGRSAR 511
Db 410 GDISSGAVSGLQLL-RPKISLFPYSTP-HPAVFICSSATPFGVGHGSGHNAK 463

RESULT 10
S09189
49K protein - Streptomyces lividans
C:Species: Streptomyces lividans
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Oct-1999
C:Accession: S09189
R:Burnett, W.V.; Henner, J.; Eckhardt, T.
Nucleic Acids Res. 15, 3926, 1987
A:Title: The nucleotide sequence of the gene coding for Xp55, a major secreted protein f
A:Reference number: S07386; MUID:87231086; PMID:3453116
A:Accession: S09189
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-469 <BUR>
A:Cross-references: EMBL:Y00142; NID:G47218; PIDN:CAA68336.1; PID:G47219
A>Note: this sequence was submitted to the EMBL Data Library, May 1987
C:Superfamily: phytoene dehydrogenase

Query Match 11.8%; Score 327; DB 2; Length 469;
Best Local Similarity 27.6%; Pred. No. 7.3e-15;
Matches 147; Conservative 64; Mismatches 223; Indels 98; Gaps 25;

5 LDAVVVGSGHNAVSAAYLAREGMSVEVLEKDTVLGGAVSTVE-RFPYKVDGSSAHL 63
2 LDAVVVGAGPNGLTAVALARRGPPVAFVFAQGTGGGARTTELTLPGRHDP 61
64 IRHSGIIEELGAGGLRYIDCDPWAPAPAGTGDGPGIVFHRDLATCSIERACGTD 123
62 GINSFAPRGLFLERYGLEWHL-PGLPMAHPFP--DGSAAVLSRVGETAASF----GARD 114
124 ADAYR---RFVAVMSRSRHVKAFSTPTGSMNLIAGFGLATARGNSLSRQFLA--P 177
115 AGPYRLIERFLPNDTLARDPM---SLPIT-----ALPRDPTVLARFGLVCLP 160
178 GDALLDEYFDSALKALAWFGAQSGPPMSEPGTAPMVGPAALMEVLPGRHVGSGALS 237
161 PSTMLRRFRDEKAKTLFAGLVAVH-----MAPLGGFAT-----GAIGLVFALA 204
238 AALASMAVDGATVALGCGVTSIRNSNHWTTTSGREVHARKVI-----AGCHIL 289
205 AHARGWPVARGSSQSSISDALTYLK-----DLGAVHTDYEVRKLDLDPARAYVL 255
290 TT-----LDLNGGDFRTTLDHWRKIRVPGICGAVLRLATSLPSVRGDATTRETS- 343
256 DTSPTALARIAGLG-----SHY-ANRYGPSVKIDYALDGPVP-----WTAEPRSA 302
344 GLQLLVSDRAHLRTA-HGAALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQ 402
303 GTVQIGADSTGTEIGALHAASOTDRAPERPFLLITVQPSVADFTRAPAGKH--VF 356
403 RLSGH--RDNASVABEADRVIGMEAPAGFTSDVLDRTIOTPRDIES-ELGMIGNM 459
357 ---GHVPGNMGDL---TDAMERQLERFAPGRDRLARATAGPAELAAARNANTVGGDIS 410

```


[illegible]

RESULT 13

RESULT 12

F70782

Probable oxidoreductase - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: F70782

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70782

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-535 <COL>

A:Cross-references: GB:273101; GB:AL123456; NID:93261565; PIDN:CAA97372.1; PID:gl314036

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: Rv0897c

C:Superfamily: phytoene dehydrogenase

Query Match	10.9%; Score 303; DB 2; Length 535;
Best local Similarity	24.4%; Pred. No. 3.8e-13;
Matches 147; Conservative	74; Mismatches 214; Indels 168; Gaps 25;
Qy	6 DAVVGGSHALYSAAYLAREGMSVELEKDTVLGGAVSTVERFPGYKVDKRGSSAHLM-I 64
Db	9 DVVVVGGHGLVAAYLAEAGLRVLLERLAQTGGAASIQAFDGVVEALSRYSLVSL 68
Qy	65 RHSGIIEELGHLGHLRYIDCDPWAPAPAGPTDG-PGIVFHRDLATCQSIERACGTD 123
Db	69 LPSRIIVADIGAPVRLAR----RPFSSYTPAPATAGRSGLLIGTGPRAHLAAIGAAPD 124
Qy	124 ADAYRRF-----VAVV-----SERSRH-----VMKAFSTPTPTGSNL 154
Db	125 AHGFAFYRRCRLVTLRWPTLLEPLRTREQARRDIVEYGGHEAAAAQWAMWDEPIGHA 184
Qy	155 IGAPGGLATARGNSHLSROFLAPGDALLDEYDFSEALKKAALWFGAQSGPXPSEPGTAPM 214
Db	185 AGAVA-----NDLRRGVIAI-DALIGTF-----ARMHEPSLMQN 217
Qy	215 VGFAALMHVLPPGRAV-----GGSGALSALASRMADVGGATVALG-----DGVTSIR 261
Db	218 ICP-LYHLVGGGTGVWHVPVIGGSGVTSALATAARHGAEIVTGADVFDALPDG--TVR 273
Qy	262 RNSNHVTVTTESGRE--VHARKVIACHLITLUDLLGNGGFDRTTLDHWRKRTVRPGIG 319
Db	274 YHSD-----GSDGAEHLVRGRFVLGVTVPAVLASLIG-----BFAALAFGAQ 316
Qy	320 AVLRLATSALPSYRGDATT-RESTSGILQLLVSDRAHLRTAHGAALAGELPERRPAVLGMSF 378
Db	317 VKNVVRVRLRLRDSVTPQQAFGTFHVNWTSQLDAAYSQAASGRLPDPLPCRAYCH 376
Qy	379 SGIDPTI-----APAGRHOVTLMSOWQYRLSGHRDWA SVAEAEADRI VGEMEAFAPGFT 433
Db	377 SLTPSPILSARLRDACAQTLTVPLGHTPESVFGD-----TEGLARLTAAVLA----SL 426
Qy	434 DSVLDRFIO-----TPRIESRLGMIGNMVMEVSLDQMWLRPLPEL 477
Db	427 NSVLAEPIDQLWTDQAQSPCIETTTTLDQRLTGMTGGNIFGALS-----W----- 474
Qy	478 SGHRVFGADG-----LYLTGASTHPGGVSGAGSGSAAARIALS--D 516
Db	475 -----PPADNDPLDTPARQWGVATDHERIMLCGSGARRGAVSGIGGHNAAVLACLA 529
Qy	517 SRR 519
Db	530 SRR 532

RESULT 13

70915

Hypothetical protein Rv1432 - Mycobacterium tuberculosis (strain H37RV)

Species: Mycobacterium tuberculosis

Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

Accession: C70915

Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

tature 393, 537-544, 1998

Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

Reference number: A70500; MUID:98295987; PMID:9634230

Accession: C70915

Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA

Residues: 1-473 <COL>

Cross-references: GB:Z95844; GB:AL123456; NID:g3250713; PIDN:CAB09252.1; PID:g2131064

Experimental source: strain H37RV

Genetics: Rv1432

Superfamily: phytoene dehydrogenase

Query Match

Best Local Similarity 10.3%; Score 284; DB 2; Length 473;

Matches 139; Conservative 77; Mismatches 202; Indels 154; Gaps 25;

7 AVVVGSGHNLVSAAYLAREGMSVEVLE-KDTVLGGAVSTVFRPGYKVDGRGSAHLMIR 65

4 AVVVGSGHNLVSAAYLAREGMSVEVLE-KDTVLGGAVSTVFRPGYKVDGRGSAHLMIR 63

66 HSGIIEELGLGAHGLRY----ICDPMWAFAPPTGPGIVFHRDLDAQCQSI----- 115

64 GSPFWAAIDLQRYGLTWKWPVDC-----AHLDDGTAGVLYRSIEATAAGLGPDKR 116

116 -ERAGCTKADAYRRFVAVWMSRSHVWKAFTPTGNSNLIGAFGLATARGNSLSRQF 174

117 WORVGD-----LAAGFDELAEDLRLVLMWPRPIRLARFGPRAALPATAMARR-- 166

175 LAPGDALIDYDFDSEALKAAALWFAQSGPPMSBPQTAP---MVGFPAALMHVLPQGRAVG 231

167 -----FHTERARALFGGAAAHVYTLDFPLTASLGLMLASGHRHGWFPAR--G 213

232 GSGALSAALARMVAGTATVAGCGVTSIRNSNHWVTWTTESGREHARKVIAAGHILTT 291

214 GSGSITKALAAALDAYGGTVATG-----VTVTSRRDIPDADIV-----M 252

292 LDL-----LG-MGGFDRTLDHWRKIRVGPVIGAVLRLATSAL-----PS 331

253 LDLSPAVALGIYGVDMPTIRNSTRFRAG-----SSAFKVDFALEGDVGVNTNPD 302

332 YRGDATTRESTSGQLLVSDRAHLRTAAGLAALAGELPPPAVL-----GMSFSGID 382

303 CRRAGTVHLGGTFABEADTERQRAQ-----GTMVQRFFVLVGGQYLADPSRSVGNIN 354

393 PTIAPAGHQVTLNSQWQPYRLSCHRDWASVAEADRIVGEAEAPGFTDVLDRPIQ 442

355 P-----IWAY-----AHVPFGTGTGATA-AVIDQIERFAPGR-----DRIVA 391

443 TPRDIESEL-----GMIGGNVHWVMSLDQM-MLWRPLPELSGHRVPGADGLYLTGASTH 496

392 TVSTSTTELYQYNNFEGGDI--IGGANDRLQVIFRRAVDVPAI-GVPGVYLCSQSAP 448

497 PGGVSGASGASAAIALSDSRKASQWRR 528

449 PGAGIHGLGCGYAAESAL-----RWLRK 471

SULT 14

0854

Hypothetical protein Rv2997 - Mycobacterium tuberculosis (strain H37RV)

Species: Mycobacterium tuberculosis

Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

Accession: H70854

Accession: H70854

R; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A; Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A; Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A; Reference number: A70500; MUID:98295987; PMID:9634230

A; Accession: H70854

A; Status: preliminary; nucleic acid sequence not shown; translation not shown

A; Molecule type: DNA

A; Residues: 1-480 <COL>

A; Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAAL6082.1; PID:g279159

A; Experimental source: strain H37RV

C; Genetics:

A; Gene: Rv2997

C; Superfamily: phytoene dehydrogenase

Query Match

Best Local Similarity 10.2%; Score 283; DB 2; Length 480;

Matches 136; Conservative 65; Mismatches 241; Indels 88; Gaps 20;

QY 5 LDAVVVGSGHNLVSAAYLAREGMSVEVLEKDTVLGGAVSTVE--RFPYKVDGRGSAHL 62

Db 1 MDVTVVGSGHNLVSAAYLAREGMSVEVLEKDTVLGGAVSTVE--RFPYKVDGRGSAHL 60

QY 63 MIRSIIIEELGLGAHGLRYDCDPMWAFAPPTGPGIVFHRDLDAQCQSIERACGFK 122

Db 61 LALASFPFAEYDLEFARGVT-LFVPDIAYANPLGR--PAATAYHDLAHTCAKLD----- 111

QY 123 DADAYRRFVAVWMSRSHVWKAFTPTGNSNLIGAFGLATARGN-SELSRQFLAPGD-- 179

Db 112 DGASWRRLGLPLVAHSEVTVVEF-----MLSDKRSLLPTALGSLVLRGLRMLAQGTPA 162

QY 180 --ALLDEYFDSSEALKAAALWFAQSGPPMSBPQTAPMVGFPAALMHVLPQGRAVCSGALS 237

Db 163 WRSAGE--DARALFTGVAHAISPLPSLVSAAGLML--ATLAHSVGNPIPVGCTQAI 218

QY 238 AALASRMVADCATVALGCGVTSIRNSNHWVTWTTESGREHARKVIAAGHILTLDDLGN 297

Db 219 DALLADLRAHGRLAAGVEITEPORSVVVFDATALLRVY----- 259

QY 298 GGFORTLIDH---WRRKIRVGPVIGAVLRLATSAL---PSYRGDATTRESTSGQLLV 349

Db 260 ---RDKLPHRYAKALRRYFRAGIAKVDVFLSDIPEMSDPLRRAATLH-----LG 307

QY 350 SDRHLRTAAGLAALAGELPPPAVLGMSFSGIDPTIAPAGHQVTLNSQWQPYRLSCHRD 409

Db 308 GTRDMARAEADVAAGRHADWPMVLAAC-----PHVADPGRIDET---GRPPFTYAHVP 359

QY 410 WASVAEAEADRIVGEAEAPGFTDVL-DRFIQTTPRDIESELGMICGNVHWVMSLDQM 468

Db 360 SGSTLDA-TETVTSVLERFAPGFRDIVVAARAVFAARMADHNANYVGGDI-----TVGAN 413

QY 469 MLWR-----PLPELSGHRVPGADGLYLTGASTHPCGGYSGASGRSNAIAL 514

Db 414 STWRALAGTPRLNPWRTP-IPKRYLCSAATPPGAVHGMCGMYAARTLL 462

RESULT 15

T36968

probable phytoene dehydrogenase (phytoene desaturase) - Streptomyces coelicolor

C; Species: Streptomyces coelicolor

C; Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C; Accession: T36968

R; Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1999

A; Reference number: 221607

A; Accession: T36968

A; Status: preliminary; translated from GB/EMBL/DBJ

A; Molecule type: DNA

A; Residues: 1-523 <SEE>

A; Cross-references: EMBL:AL109962; PIDN:CAB53153.1; GSPDB:GN00070; SCOPDB:SCJL.35

A; Experimental source: strain A3(2)

```

;Genetics:
;Gene: crtB; SCOREDB:SCU1.35
;Superfamily: phytoene dehydrogenase

Query Match      9.9%; Score 275; DB 2; Length 523;
Best Local Similarity 25.9%; Pred No 3e-11;
Matches 150; Conservative 65; Mismatches 223; Indels 142; Gaps 26;

y 8 VVVGSGHNLVSAAYLAREGMSVEVLEKOTVLGAVSTVERPPGYKVDGSSAHLMIKHS 67
b 12 VVVGAGLGLACALHLLGAGRRVTWVERDAGPGGRSGRV-RLGGYELDTGTV-LTMDPL 69
y 68 GIIEELGLGAGSLRYIDCDPNAFPAPGDTGPGIVFHRDLATQCSIERACGTXDADAY 127
b 70 ADEAFAGVGSLSLRREVELTALDPAYACFADGSGALDVHTDGEAMEAEVRFAGPAQAGY 129
y 128 RRFVAVMSER-----SRHVMKAFSTP-----PTGSNL--IGAPGGLATARGNSLSRQF 174
b 130 RD-LRRWLESLYRAQMRRFIDTNFDSPLQLLHPDLARLAALGGFGRLDGRIG-----RF 182
y 175 LAPGDALLDEYFDSEALKAAALANFQAQSGPPMSEPGTAPMVGFAALMHV-----LPPG 227
b 183 LS--DERLRVVFQALYAGVA-----PARA-LAAYAVIAYMDTVAGVWFFK- 226
y 228 RAVGGSGALSAAALASRMADVGTALGDGVTSTIRNSNH-WTVTTESGREVEHARKVIAGC 286
b 227 --CGMHALPRAVADAAATAGADLRWSAEVKALERSAGRVRVAVHLASGERIACDAVLTC 283
y 287 HILTTLLGNGGFDRTLLDHWRRKIRVGPICGAVLRATSLALPSYRGDATTRESTGLQ 346
b 284 ELSTAYGLLGR-----APRRPARLRHSPSAVILHAG----- 314
y 347 LLVSDR-----AHLRTAHGAA-----LAGELPPRAVLGMSFGIDPTIAPAGR- 391
b 315 ---TDRTWPHLAHHTLSFGAAWERTPEELTRYGELMSDFSLLITRTHDPALAPGRHL 371
y 392 -----QVTLMSQMOP-YELSGHRDMSVAEAEADRIVGENEAFAPGFTDS 435
b 372 EYVLAPCPNTDVGPDVAWVRDLGPRYR-----ESLVGVLESRGLEGFADS 416
y 436 VLDRFIOTPRDISSELGICGMVYHVMESLDQMWLRPLPELSGHRVPGADGLYLTGAST 495
b 417 VQELLVTPLDWDAQ-GHAAGSPFSVSHTFACQTGPRP-----RNLYRGLDNVVLACGT 470
y 496 HPGGVSG--ASGR-SAARIALSDSRGKASQMMRRSRSS 532
b 471 TPGVGVPVLVSGKLAARVT-----CGAGSRPARTRRS 504
```

Search completed: February 29, 2004, 14:53:00
Job time : 17.808 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:51:24 ; Search time 35.3218 Seconds

(without alignments)

3180.293 Million cell updates/sec

tie: US-09-941-947A-38

fect score: 2768

quence: 1 MSFIDAVVCSGHNALVSA.....ALSDRRKGASQMRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 211153259 residues

tal number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2768	100.0	532	10	US-09-941-947A-38
2	2764	99.9	532	14	US-10-209-372-2
3	859.5	31.1	511	14	US-10-209-372-4
4	859.5	31.1	511	15	US-10-369-493-23397
5	888	24.9	542	14	US-10-209-372-6
6	888	24.9	542	15	US-10-369-493-2831
7	660.5	23.9	539	15	US-10-369-493-18878
8	651.5	23.5	537	15	US-10-369-493-19986
9	537.5	19.4	539	15	US-10-369-493-12005
10	484	17.5	521	15	US-10-369-493-10809
11	477	17.2	518	15	US-10-369-493-20413
12	471.5	17.0	517	15	US-10-369-493-4938
13	471.5	17.0	517	15	US-10-369-493-7696
14	469	16.9	524	15	US-10-369-493-12144
15	466.5	16.9	544	15	US-10-369-493-5969

16 453.5 16.4 512 15 US-10-369-493-4801
17 453.5 16.4 512 15 US-10-369-493-7560
18 443 16.0 544 14 US-10-156-761-8540
19 424 15.3 501 15 US-10-369-493-10498
20 363.5 13.1 523 15 US-10-369-493-12145
21 331 12.0 472 14 US-10-156-761-9474
22 309 11.2 533 14 US-10-156-761-8803
23 296.5 10.7 471 9 US-09-738-626-4041
24 293.5 10.6 471 14 US-10-156-761-8883
25 270 9.8 501 15 US-10-369-493-28850
26 261 9.4 513 14 US-10-156-761-8562
27 249 9.0 512 15 US-10-369-493-18644
28 248 9.0 506 15 US-10-369-493-8142
29 247 8.9 496 15 US-10-369-493-20440
30 247 8.9 536 15 US-10-369-493-20048
31 229.5 8.3 494 15 US-10-369-493-20907
32 226 8.2 511 15 US-10-369-493-20907
33 225 8.1 495 15 US-10-369-493-7895
34 225 8.1 497 14 US-10-156-761-13629
35 219.5 7.9 548 9 US-09-738-626-4194
36 217.5 7.9 494 9 US-09-547-267-5
37 217.5 7.9 494 10 US-09-920-923-4
38 213.5 7.7 491 15 US-10-369-493-234
39 212.5 7.7 498 15 US-10-369-493-19248
40 209 7.6 498 15 US-10-369-493-20438
41 208.5 7.5 506 15 US-10-369-493-17830
42 203.5 7.4 518 15 US-10-369-493-7750
43 200 7.2 485 15 US-10-369-493-19518
44 197.5 7.1 548 15 US-10-369-493-540
45 192.5 7.0 530 14 US-10-128-713A-18

ALIGNMENTS

RESULT 1

US-09-941-947A-38
; Sequence 38, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: Dicosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odum, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis AN12
US-09-941-947A-38

Query Match 100.0%; Score 2768; DB 10; Length 532;
Best Local Similarity 100.0%; Pred. No. 3.6e-238;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIDAVVCSGHNALVSAAYLAREGWSVEVLEKDTVLGAVSTVERFPGYKVDGRSSA 60

DB 1 MSFIDAVVCSGHNALVSAAYLAREGWSVEVLEKDTVLGAVSTVERFPGYKVDGRSSA 60

QY 61 HLMTRHSGIIEELGLGAGRLRYIDCDPFAFPAPGTDGPGIVFHRDLDTATCSIERACG 120

347 GYGEYLAQPTTDFPLVAMSFSAVDDSLAPENGVDVLMWAQYYPFELA-TGSEWETRAEA 405
418 ADRIVGEMEAFAFGTOSVLDRTIOTPRDIESLGIMTGGNVHVMVEMSLDOMLWRPLPEL 477
406 RENILRAFEHVPAGTROTIVGELVOTQWLETNLGLHGRNVHLEKSFQDMFSFRWLKA 465
478 SGHRVPGADGLYLITGASTHPGGVSGSGSRAARIALSDSR 519
466 SQYRWPGVQGLYLITGASTHPGGGIMGASGRNAARVIVKDLTR 507

SULT 4

-10-369-493-23397
Sequence 23397, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 23397
LENGTH: 511
TYPE: PRT
ORGANISM: Deinococcus radiodurans
-10-369-493-23397

Query Match 31.1%; Score 859.5; DB 15; Length 511;
Best Local Similarity 39.7%; Pred. No. 6.8e-68;
Matches 207; Conservative 80; Mismatches 208; Indels 27; Gaps 11;

6 DAVVVGSHNALVSAAYLAREGNSVEVLEKDTVLGGAVSTVERFGYKVDGRGSSAHLMR 65
5 DLIWVGAGHNLVTAAYAAAGLKVGFERRHIVGGAVSTEEVVPGRFDYGGSAHLIR 64
66 HSGIIEELGLGAHGLRYIDCDPWAFAPAGTDPGIVF-HRDLATCQSI-ERAQTKD 123
65 MTPIVRELELTHGLHYLEVDPMFHA-----SDGETPWFHIRDAGTIRELDEKPG--Q 117
124 ADAYRRFVAVNSERSHVMKAPSTPTGSGNLIGAPGLATARG-----NSELRSQPLAPG 178
118 GDAYGRFLDDMTPPARAVADLFNSAPGLDL----GQWVRSGQKDMNEQLPR-TLRPY 172
179 DALDDEYFDSEALKAAALWAFGAQSGPPMSBFGTAPMVGFALMHVLPFGRAVGGSGALSA 238
173 GDVAREYFSEERVAPLTWAAQSGPPSPDPLSAPFLMHPLVHEGGVAPKGGSGGLTK 232
239 ALASRMVADGATVALGQVTSIR-RNSNHWVTTEGSRVHARKVJAGCHILITLDDLGN 297
233 ALRATAREGGEVFTDAPVKEILVQKQACIRLESGETYTARAVSGVHILITANAL-- 290
298 GGPRTTLTDHWRKIRVGPGLGAVLRATLSALPSYRGDAITTESTSGLOLLVSDRAHLRT 357
291 ---PAEVPSAARNVRVNGVGFMLRLALSEKVKYR-HTEPDSR:GLGLLNKNERQIMQ 346
358 AHGAALAGEPLPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYLSGHRDMSVAEAE 417
347 GYGEYLAQPTTDFPLVAMSFSAVDDSLAPENGVDVLMWAQYYPFELA-TGSEWETRAEA 405
418 ADRIVGEMEAFAFGTOSVLDRTIOTPRDIESLGIMTGGNVHVMVEMSLDOMLWRPLPEL 477
406 RENILRAFEHVPAGTROTIVGELVOTQWLETNLGLHGRNVHLEKSFQDMFSFRWLKA 465
478 SGHRVPGADGLYLITGASTHPGGVSGSGSRAARIALSDSR 519

Db 466 SQYRWPGVQGLYLITGASTHPGGGIMGASGRNAARVIVKDLTR 507
RESULT 5
US-10-209-372-6
Sequence 6, Application US/10209372
Publication No. US20030100045A1
GENERAL INFORMATION:
APPLICANT: E. I. du Pont de Nemours, Inc.
APPLICANT: Cheng, Qiong
APPLICANT: Zao, Luan
TITLE OF INVENTION: CAROTENOID KETOLASE GENE
FILE REFERENCE: CL-1849 US NA
CURRENT APPLICATION NUMBER: US/10/209,372
CURRENT FILING DATE: 2002-07-30
PRIOR APPLICATION NUMBER:
PRIOR FILING DATE:
NUMBER OF SEQ ID NOS: 47
SOFTWARE: Microsoft Office 97
SEQ ID NO 6
LENGTH: 542
TYPE: PRT
ORGANISM: Synecocystis sp. PCC6803
US-10-209-372-6

Query Match 24.9%; Score 688; DB 14; Length 542;
Best Local Similarity 31.9%; Pred. No. 1.5e-52;
Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

QY 6 DAVVVGSHNALVSAAYLAREGNSVEVLEKDTVLGGAVST-----VERFPQYKVDGRGSSA 61
Db 5 DWIIVGAGHNLVCAAVLQGLGVLTLEKREVPGGAATTEALMPELSPQFRNRCALDH 64
QY 62 LMIRHSGIIEELGLGAHGLRYIDCDPWAFAPAGTDPGIVFHRDLATCQSIERACGT 121
Db 65 EFIFLGPVLQELNLAQYLEYLFCDPSVF---CHGLDQAFMSYRSLEKTCAMH-ATYSP 120
QY 122 KDADAYRRFVAVNSERSHVMKAPSTPTG-----SNLIGAPGLATARGNSEL 171
Db 121 RDAEKYQFVNYWTDLLNAVQPAFNAFPQALLDLALMYGWNLSVLAIAGSKTKALDFI 180
QY 172 RQFLAPGDALLDEYFDSEALKAAALWAFGAQSGPPMSBFGTAPMVGFALMHVLPFGRAVG 231
Db 181 RTMTGSPEDVLNWFDSERVKAPLARLCSIGAPPQSGSSGGMVMMVHEHLEGIARPKG 240
QY 232 GSGALSAAALASRMVADGATVALGQVTSIR-RNSNHWVTTEGSRVHARKVJAGCHILT 290
Db 241 GTGALTALVKLVQAQCKILITQTVKSVLVNQNQAIGVEVANGEQYRCK-----GVIS 295
QY 291 TLD-----LLNGSGFDRFTLDHWR-KIRVGFPGICAVLRL--ATSALPSYRGDAITRE 340
Db 296 NIDARLFLQLIVERGALAKVQNGLGERLERRTVNNNEAILKIDCALSGLPHFTAMAGP-E 354
QY 341 STSGIQLLVSDRAHLRTAHGAALAGELPPRAVLGMSF-SGIDPTIAPAGRHQVTLWSQ- 398
Db 355 DLTUTILIADSVRHEERAHALIALGQIPDANPSLYLDIPTVLDPTWAPGQH--TLWIEF 412
QY 399 WQPYRLSG-----HRDWA-SVAEAEADRIVGEMEAFAFGTDSVLDRTIOTPRDIESE 450
Db 413 PAPYRIAGLEGTGLMGGTWTDDELKEKVADRVIDKLTDPAPNLKSLIIGRRVESPALAQ 472
QY 451 LGMIGGNVHVMVEMSLDOMLWRPLPELSGHRVPGADGLYLITGASTHPGGVSGSGSAA 510
Db 473 LGSYNGNVYHLDMSLDQMFLRPLPELIANYQTP-INKLYLTGAGTHPGGSGISGPNRCA 531
QY 511 RIALSDSR 519
Db 532 RVFLKQQR 540

RESULT 6
US-10-369-493-2831

Sequence 2831, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 2831

LENGTH: 542

TYPE: PRT

ORGANISM: Synechocystis sp.

US-10-369-493-2831

Query Match 24.9%; Score 688; DB 15; Length 542;

Best Local Similarity 31.9%; Pred. No. 1.5e-52;

Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

6 DAVVGSNNALVSAAYLAREGWSVEVLEKDTVLGAVST-----VERPGKYVDRGSSAH 61

5 DVVVGAGHGLVCAAYLQRLGVTLLKREVPVGGAAATTEALNPESLPQFRFNCAIDH 64

62 LMRHSGIIEELGAGHGLRYIDCDPAFAPAPGTDGPGIVFHRDLDTQCSIRACGT 121

65 EFTLPGVLOELNLAQYGLYLEFCDSVP---CPGLDGAQFMSYSLKTCACAH-ATYSP 120

122 KDADAYRRFVAVMSERSRHVKAFSTPTTG-----SNLIGAFGLATARGNSLS 171

121 RDAEKYRQFVNYWTDLLNAVQPAFNAAPPQALLDLALNYCWNLSKSLVALAGSKTKALDFI 180

172 RQFLAPGDALLDEYFDSSEALKALAWFGAQSGPPMSEPGTAPMVGFALMHVLPGRVAVG 231

181 RTWIGSPEDVINEFWFDSERVKAPLARLCEIGAPPSQKSSGMMVAVRHLEGIARPKG 240

232 GSGALSAALASRMVAVGATVALGDGVTISR-RNSNHWTVTTSGREVHARKVIACHILT 290

241 GTGALTEALVKLVAQGGKILDTQTVKRVLVENNOALGVEVANGEGYRAKK-----GVIS 295

291 TLD-----LLNGGDFRTTLDHWR-KIRVGPGLGAVLR--ATSALPSYRGDAITRE 340

296 NIDARELFLOLVBPGLAKVQNGLERLRTVNNNEALIKIDCALUSGLPHTNAGP-E 354

341 STSGLOQLVSDRAHLRTAHGAALAGELPPRPVAVLGMSP-SGIDPTTIAPAGRHQVTLMSQ- 398

355 DLCTGTLIADSVRVEEAHALIALGQIPDANPSLYLDIPTVLDPQTPWAPQGH--TLWIEF 412

399 WQPYRLSG-----HEDWA-SVAEAEADRIVGEAEAFAPGPTDSVLDRIPTPDISE 450

413 FAPYRIAGLEGTGLMGSGTDELKRVADRVIDKLTIDYAPNLSKLIIGRVSSPAELAQ 472

451 LGMIGGNVHVMESLDOMLWRPLPELSGRVPGADGLYLTGASTHPGGGVSGASGRSAA 510

473 LGSYNGVNYHLDMSLDQWMLRPLPEIANVQTP-INKLYLTGAGTHPGGSISSMPGRNCA 531

511 RIALSDSR 519

532 RVFLKQQR 540

RESULT 7

US-10-369-493-18878

Sequence 18878, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 18878

LENGTH: 539

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-10-369-493-18878

Query Match 23.9%; Score 660.5; DB 15; Length 539;

Best Local Similarity 31.7%; Pred. No. 4.2e-50;

Matches 174; Conservative 94; Mismatches 224; Indels 57; Gaps 17;

6 DAVVGSNNALVSAAYLAREGWSVEVLEKDTVLGAVSTV-----RPGKYVDRGSSAH 61

5 DVVVGAGHGLVCAAYLQRLGVTLLKREVPVGGAAATTECLPKRAPGKFNLCALDH 64

62 LMRHSGIIEELGAGHGLRYIDCDPAFAPAPGTDGPGIVFHRDLDTQCSIRACGT 121

65 EFTLPGVLOELNLAQYGLYLEFCDSVP---CPHDPGKYFLAKSLEKTCACAEIAR-YSE 120

122 KDADAYRRFVAVMSERSRHVKAFSTPTT-----TGSNLIIGAFGLATARGNS-----ELS 171

121 RDAKKYAEFTYVQRAIGAMIPMENAPPKSIIDIVGNYDITKFDLFSVIGSPNKTLDPI 180

172 RQFLAPGDALLDEYFDSSEALKALAWFGAQSGPPMSEPGTAPMVGFALMHVLPGRVAVG 231

181 RNMWLTSAEDILNMFDSFELKAPLARLCEIGAPPSQKTIICALMNMARHNPCHMARPKG 240

232 GSGALSAALASRMVAVGATVALGDGVTISRNRNHWTVTTSGREVHARKVIACHGCH- 287

241 GTGALIKALWNLVSKGGVILTDQVQEK-----VLIDGKAVGVV-VSGGTEYRAK 290

288 --LLTLD-----LLNGGDFRTTLDHWR-KIRVGPGLGAVLR--LATSALPSYRGD 335

291 YGVISNIDAKRLFLQMTDKSDVDADPDLMWRLEHRIVNNNETILKIDLDLDEPLRFPFH 350

336 ATTRESTSGLOQLVSDRAHLRTAHGAALAGELPPRP-PAVLGMSFSGIDPTTIAPAGRHQVT 394

351 AKDEYLVGSILLIADSVHAEQAHSKCTLGEIPDSDESVMVWVPSYLDPTLAPSGKH--T 408

395 LMSQ-WQPYRLSGH-----DNASVAEAE-ADRIVGEAEAFAPGPTDSVLDRIPTQTR 445

409 VWIEFTAPYQIAGAEQTFKGTGMDTDLKQVADKYVDKLATYAPNVKTATLARVESPA 468

446 DIESELGMIGNVHVMESLDOMLWRPLPELSGRVPGADGLYLTGASTHPGGGVSGAS 505

469 ELGERLCAKNGVNYHLDMSLDQWMLRPLPEIANVQTP-IDNLYLTGAGTHPGGSISSMP 527

506 GRSARIAL 514

528 GRNCARVFL 536

RESULT 8

US-10-369-493-19986

Sequence 19986, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 2002-02-21
LENGTH: 537
TYPE: PRT
ORGANISM: No. US20030233675Altoc punctiforme
-10-369-493-19986

Query Match 23.5%; Score 651.5; DB 15; Length 537;
Best Local Similarity 32.0%; Pred. No. 2.7e-49;
Matches 176; Conservative 90; Mismatches 225; Indels 59; Gaps 17;

6 DAVVVGSHNALVSAAYLAREGMSVEVLEKDTVLGGAVSTVE----RFGYKVDGSSAH 61

5 DVVLIGAGHGLVCAAVLLKAGSVLLLEKRSVPGGNAITEELCPQEAQGFKNLCAIDH 64

62 LMIHSGIIIEELGAGHGLYIDCDPWAPAPAGTDPGIVFHRDLDTQCSIERACGT 121

65 EPIHLPVWSELEKYLHLECDPVVF---CPHPDKYFLGHKSLEKTCABEIAF-YNE 120

122 KQADYRRFVAVMSERSRHVKAFSTPP-----TGSNLIGAFGLATARGNS---ELS 171

121 RDAKYAEFVDYWRICAMIPFNADPKSIIDIVGNYDIKKFKDLSFVIGSPNKTLDPI 180

172 RQFLAPGDALLDEYFDSALKALAWFQAQSGPPMSEPGTAPVMVGFPAALMHVLPGRV 231

181 RTMLTSAEDLLNEWFDEEFLKAPLARLASSELGAPPSQKTLAIGAIWMMRHNPGMARPG 240

232 GSGALSALASRMADGATVALGCVTSIRNSNHWTVTTESGREVHARKVLAG----- 285

241 GTGALVQALVNLVTSKGVILTQCHVEK-----VLIDDGKAVGVR--VAGGKEYRA 289

286 -CHILITLD-----LLGNNGGFRRTTLDHWR--KIRVGPVIGAVLR--LATSALEPSYR 334

290 KYGVISNIDAKFLQMTDSDVDGADPDLWERLERIVNNETILKIDLALDEPLHFFH 349

335 DATTRESTGQLLVSDRAHLRTAHGAALAGELPPR--PAVLGMSFGSIDPTIAPGRHGV 393

350 HAHKDEYLVGSLIADSVAHQASHKCTLGEIPDADPSMYVMPSYLDTLAPPCKH-- 407

394 TLWSQW-OPYRLSGR-----DWA-SVAEAEADRVIGEMEAFAFGFTDSVLDRFIQT 444

408 TWLIEYAPYQIAGAGTGLKGTWTDLKNKYADRVVDKLDADYAPNVMATIAARVESF 467

445 KDISELGMIGNNVHMVMSLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSGA 504

468 AELGERLGAVKGYHYVDWTDQIMFFRPLPEIANYKTP-IDNLFITGAGTHFGGSISQW 526

505 SGRSAARIAL 514

527 PGRKCARAFI 536

RESULT 9
-10-369-493-12005
Sequence 12005, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12005
LENGTH: 539
TYPE: PRT
ORGANISM: Mesorhizobium loti
US-10-369-493-12005

Query Match 19.4%; Score 537.5; DB 15; Length 539;
Best Local Similarity 30.4%; Pred. No. 4e-39;
Matches 168; Conservative 92; Mismatches 218; Indels 75; Gaps 21;

Qy 6 DAVVVGSHNALVSAAYLAREGMSVEVLEKDTVLGGAVSTVERFPGYKVDGSSAHLMI 65

Db 5 DAVVVGSHNALVSAAYLAREGMSVEVLEKDTVLGGAVSTVERFPGYKVDGSSAHLMI 64

Qy 66 HSGIIEELGAGHGLYIDCDPWAPAPAGTDPGIVFHRDLDTQCSIERACGTQAD 125

Db 65 P3-IMRDLPLPRFGLQVISEGGAVFT---RDGDLANYRDHDAHRRBP-FSRRDAE 118

Qy 126 AYRRFVAVMSERSRHVK-AFSTPTGSNL---IG-----AFGLATARGNSLSRQ 173

Db 119 AYRYARDVTQCRFIQFLMRTPADPTSPKPDIGELLYLGGKFAGLSAEAKLTURFW 178

Qy 174 FLAPGDALLDEYFDSALKALAW---FGAQSPPMSEPGTAPVMVGFPAALMHVLP 226

Db 179 TMSISD-FLDEYPTDVIKANFALSIGITAGP--MSPGTA---YVLLHYMGEVDGS 231

Qy 227 ---GRAVGGSALSALASRMADGATVALG---DQVTSIRNSNHWTVTTESGREVHA 279

Db 232 VGWGYARGGMAVTKAASFKASGGTIRTGAEDHVLVSRGKAK--GVVLGGBBEYIG 289

Qy 280 RKTAGACHILTLDLGNGGDFRTTLDHWR-----KIRVGPVIGAVLRATLALPS 331

Db 290 KLVSNADVKTFKLVE---EKELPDIPLRRVKNFKIR---GSSGKVNIADLSLPEFA 343

Qy 332 ---YRDATTRESTGQLLVSDRAHLRTAHGAALAGELPPRPAVLGMSFGSIDPT 384

Db 344 LAXDSPVYRGDMHFTDSIERMERAYDDMK-----ACRMSADPFLDMVIFTLDPT 393

Qy 385 IAPAGRCQVTLWSQYRLSCHRDNASV-AEAEADRVIGEMEAFAFGFTDSVLDRFIQT 443

Db 394 MAPGKHFMSCFVQYAPPKVNG-RDWTADRDGPASVVAQIAEYSPGFRDRIHMEVRT 452

Qy 444 PRDIESELGTMGGVNMVHMVMSLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSG 503

Db 453 PREIEAEVGLTEGNIHQBELTFDQLLFNRVPVGAQVRESF-VGGLYMCSSSTHPGGVVG 511

Qy 504 ASGRSAARIALSD 516

Db 512 AGRNAAABILRD 524

RESULT 10
US-10-369-493-10809
Sequence 10809, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 10809

LENGTH: 521

TYPE: PRT

ORGANISM: Sphingomonas aromaticivorans

S-10-369-493-10809

Query Match 17.5%; Score 484; DB 15; Length 521;

Best Local Similarity 30.5%; Pred. No. 2.3e-34;

Matches 164; Conservative 84; Mismatches 234; Indels 56; Gaps 19;

y 6 DAVVSGGNALVSAAYLAREGWSVEVLEKDTVLGAVSTVERFPGYKVDGSSAHLIR 65
 b 2 DALIIGGNGNLVCAYLAKAGKVKVLERRDVGGAAVTEFFHFGFNSTASTVLSLR 61
 y 66 HSGIIEELGLGAGHLYIDCDPWAPAPAGT-----DGPGVFVHRLDQTCQSIERACG 120
 b 62 PK-VIADMKLHDYGVRIERTISNPF-PEDTVLKLGGPG-----RTEAFARFS 110
 y 121 TKDADAYRRFVAVNSRSRHVMKAPS--PPTGNSNLIGAFGLAT-----ARGNSE 169
 b 111 KDAESYRYDAA-LEKVANVLRIISLQTFPNVGGGIAALKAAATQGWPIAKLIDIAQFD 169
 y 170 LSRQFLAPGDALDEYFDSALKAAALAWFCAQSG--PPMSEPTAPMVGFDAALMHVLP- 226
 b 170 LLDIPTKSARDFLDGNFEDDHVKSAG-EDAVVGNFAGVSTPGSA-----YVLLHHVFGEV 224
 y 227 -----GRAVGGSGALSAAALASMAVDGATVALGCVTISIRNSNHWI-VTTESGRVH 278
 b 225 NGKLGAWHSGVGGGAIQTAMAKACVDAGVEISLEAPSVRLVNNKAAAGVKLEGEELY 284
 y 279 ARKVIAGCHILTTLLGNGGDFRTLLDH-WRRKIR-VGPFGIGAV-LRLATSAIPSYR-- 333
 b 285 APIVAAN-----VGPMLYRQWVDASDLDFERRMKNYKTSCTFRMNVALSELPDFKVL 340
 y 334 -GDATTESTSGLQLLVSDRAHLRTHAGALAGELPPRAVLGMSFGIDPTIAPAGRHO 392
 b 341 PGKQLAEHTAGI-ILAPGMDYMDQAFIDQAQFGWKKPIVE-KIESTVDDSLAPPGQEV 399
 y 393 VTLASQPVLSCHRDWASVAEAEADRIVGEMAPAPGFTDVLORFTQTPRDISELG 452
 b 400 ASLFCQFAPQLPGRSWDDCREVALIITVNDHAPNEKASVIARQIHSPDLDERKFG 459
 y 453 MIGNVMEVMSLDQMLWRPLPELSGHRVFGADGLVLTGASTHPCGSGVSGASGRSAA 510
 b 460 LIGGDI FHGTWGLDQWAAARVPLNGDYSRP-IGLYMCGSGTHPGGVTGAPGHNA 516

RESULT 11

JS-10-369-493-20413

Sequence 20413, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 20413

LENGTH: 518

TYPE: PRT

ORGANISM: Rhodospseudomonas palustris

FEATURE:

NAME/KEY: (1)..(518)

LOCATION: (1)..(518)

OTHER INFORMATION: unsure at all Xaa locations

Query Match

Best Local Similarity

US-10-369-493-20413

Query Match 17.2%; Score 477; DB 15; Length 518;

Best Local Similarity 30.9%; Pred. No. 9.5e-34;

Matches 170; Conservative 85; Mismatches 216; Indels 80; Gaps 26;

y 6 DAVVSGGNALVSAAYLAREGWSVEVLEKDTVLGAVSTVERFPGYKVDGSSAHLIR 65
 b 1 DVLIIGAGNGNLTCAAYLARAAGLKVYVRRNVVGGAAVTQEFHFGFR-NSVAATVLSLL 59
 y 66 HSGIIEELGLGAGHLYIDCDPWAPAPAGTDPGIVFHRDLDTQCQSIERACCTKAD 125
 b 60 NPKVIADLKLHEHGLRIVERKCAQNFL-PAP--DGOYLL--TGSNTTAAASLAR-LSAADA 113
 y 126 AYRFRFVA---VMSKSRH-VMKAFSTPTGSLNLCAPG-----GLATARGNSELSR 172
 b 114 AFGFAAELETTIADVLHFLVRA---PP---NLVVQFGLPAIRSSINALETANRLRALTM 167
 y 173 Q-----FLAPGDALDEYFDSALKAAALAWFCAQSGPPMS--EPGTAPMVGFDAALMH 222
 b 168 EQQRLLDLFTCSAGEMLDARFEHDLVK-ALFGFDAIVGNVASYAAGSA-----YVMLHH 222
 y 223 VLP-----GRAVGGSGALSAAALASMAVDGATVALGCVTISIRNSNHWI-VTTES 273
 b 223 AFGVNGKGVWGRAIGMGALITSAVAAAARAGABEISAGVRELVKDRVVGVTLLDD 282
 y 274 GREVHARKVIAGCH---ILTTL---DLLGNGGDFRTLLDHWRKIRKVGPGIGAV-LRLAT 326
 b 283 GENVRAFVASNVNPKLLYTRLLAPDALPD--VRRRMQMK-----TCSGTFERNVAL 334
 y 327 SALPSYR-----GDATTESTSGLQLLVSDRAHL-RTHAGALAGELPPRAVLGMSFG 380
 b 335 SRLPSFTALPGDGDHLTAGIIIPSLGYMDRAYQDARAGNS-----REPVEMLIPST 388
 y 381 IDPTIAPAGRHQVTLNSQMPYRLSGHRDWSVAEAEADRIVGEMAPAPGFTDVLDRP 440
 b 389 LDDSLAPKGVHSAFLCOHVAPELPGASWDDHDEVDLMIAVDVRYAPGFAASVLRQ 448
 y 441 IOTPRDIESLGMIGNVMVMSLDQMLWRPLPELSGHRVFGA-DGLVLTGASTHPCG 499
 b 449 ILSPLDLEREFGLVGGGIFHGALSLNOLFAARELPQADY--FGALKGLYHAGSGAQPN 506
 y 500 GVSGASGRSAA 510
 b 507 SVSGAPGNAA 517

RESULT 12

US-10-369-493-4938

Sequence 4938, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 4938

LENGTH: 517

TYPE: PRT

ORGANISM: Burkholderia fungorum

US-10-369-493-4938

Query Match 17.0%; Score 471.5; DB 15; Length 517;

Best Local Similarity 29.6%; Pred. No. 2.9e-33;

b 60 NLSLFAGSAPHRKYANELKTQGLEFAPVADC-----FASAFP--DGRWFGVNDLEKTASR 113
Y 115 IERACGTKDADAYRRPVAVKYSRRVVKAFSTPTGSLNLIAGFGLATARGNS-----SLS 171
b 114 M-AAPSAADAATWRKLVAAFPCEAEHLFRLLGSPMSARALACTANWLARKKGVAGALDTG 172
Y 172 RQFLAPGCDALLDYPFSEALKAAALAWFGAQSPPMSEPGCTAPMGVFAALMHVLP-- 226
b 173 RLLSSPRTWLETFESPVRATLATWGHLL-----DFAPDIAAGGAVFPFYLESMANQS 225
Y 227 -GRAVGSAG--LSAALASRMVADGATVALGDGVTISRRNSNHWITVTTESGREVH--ARK 281
b 226 FGMVLKGGADTIIRALAGWTSAGSKIUTGAEVSEITVSNKATGVALTSGETHTATKA 285
Y 282 VIAGCHILITDILL-----GNGGDFRTDHDHRRKIRVCGFGAGVLRATSAIPSYRGA 336
b 286 VIAGVAPKALTGLLPGSGNAGFDTA-----MKFRVAPGT--MMIHLALDELPCWRAGS 339
Y 337 TTSTSTSGLLVSDRAHLRTARGAALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLW 396
b 340 ELRQ--FAYVELSPSLDANRT--YQAWAGMLPDEPVLVVGQPTAIDPSRAPQGRH--VLM 395
Y 397 SQWQ--PYRLSGH-----RDWASVAAEADRIVGEAEAFAGFTDSVLDRIQTPRDI 447
b 396 VOVRMLPAEITGDAGGKIAPAHWDQKDAVERVLDIETVAPGLRSKILGRSVFSPIDL 455
Y 448 ESR-LGMIGNVAVHVMESLDQMLWRPLPELSCHRVPGADGLYLTGASTHPCGGVSGAG 506
b 456 ERENPLVGDDQVCGSHLAQNFLEFRPARGWNTP--VGNLHLTGAATWPGAGTGAASG 514

ESULT 15

S-10-369-493-5969

Sequence 5969, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cag, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 5969

LENGTH: 544

TYPE: PRT

ORGANISM: Caenorhabditis elegans

S-10-369-493-5969

Query Match 16.9%; Score 466.5; DB 15; Length 544;
Best Local Similarity 28.3%; Pred. No. 8.8e-33;
Matches 157; Conservative 92; Mismatches 231; Indels 75; Gaps 20;

Y 6 DAVVVGSGHVALSAAYLAREGWSVEVLEKDTVLGGAVSTVERFPGYKVDGSSAHLMI 65
b 16 DALLIGGCHGLTAAYLYLKAGKKVCVLERHHVVGAAVTEIIVGFRPSRASYLLSLR 75
Y 66 HSGIIEELGAGHGLYICDPAFAPPAGTGDGPGIVFHRDLATQCOSIERACGTDAD 125
b 76 -PVVMOELNLKFTGLRYKIRNPNSFTPIR--NTHSLLLGMDMAENQKEIAK-FSORDAG 131
Y 126 AYRRFVAVMSERSR-----HYMKAFST-PPTGSLNLIAGAFGLA 162
b 132 NYPKYEHFISEIVHSTFQELMDYEPDLQKPIHKLPHLYLLFKTVQPLGLRNADVFLM 191
Y 163 TARGNELSQFLAPGDALLDEYDSEALKAAALWFG--AQSGPPMSEPGTAPMVVGFAL 220

Db 192 T-----APISKIMNKFESDVLKATLGTGTVIGLAASEM-DPQT-----GYVLL 234
QY 221 MHVLP-----GRAVGSAGLSAALASRMVADGATVALGDGVTISRRNSN-HWTVT 271
Db 235 HHVIGGLDEHKGAWGVYVCGMGAVSNAIAECAKSHCAEITYTEQDVQEVLLDGNVAKGVRL 294
QY 272 ESGREYHARKVIAGCHILITDILLGNGGDFRTDLDHRRKIRVCGFGIGAV--LRLATSAL 329
Db 295 SNGKELHSLKIVMSNATPHVTFFHLVK---KESLPEFFHNNINQIDVTSPTVKINVAVKEL 351
QY 330 PSY-----RGDAATTRESTSGLQLLVSDRAHLRTAAGAAAL-----AGELPPPAVAVLGMFSFGI 381
Db 352 PNFIAKPNQGSSEPMPHQTTIEM---NCENMQVVDVMDVYKNGRYSRFPVIEWTIPSSV 408
QY 382 DPTIA-PAGRHQVTLMSQWCPYRLSHRDWASVAAE-ADRIVGEAEAFAGFTDSVLDRI 439
Db 409 DRTIVDSADGHVVLLFTQYTPFS-PKDGWTEEKTEYAKHVFSIDAYAPNFSSVIGY 467
QY 440 FIOTPRDISELQMGIMGNVHVMESLDQMLWRPLPELSCHRVPGADGLYLTGASTHPCGG 499
Db 468 DILTPDIOQTGTTGNIIPHGMSLDQLYVSRPISKNSNYSTP-IESIYLCGSGAHFEG 526
QY 500 GVSAGSRSAAIAL 514
Db 527 GVTGAPGRLSALHAL 541

Search completed: February 29, 2004, 15:28:09

Job time : 38.3218 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:35:44 ; Search time 16.1665 Seconds
(without alignments)
1698.885 Million cell updates/sec

tle: US-09-941-947A-38

fect score: 2768

quence: 1 MSAPLDAVVVGSHNALVSA.....ALSDRRGRASQMRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Issued Patents AA.*

1: /cgm2_6/ptodata/2/iaa/5A.COMB.pep.*

2: /cgm2_6/ptodata/2/iaa/5B.COMB.pep.*

3: /cgm2_6/ptodata/2/iaa/6A.COMB.pep.*

4: /cgm2_6/ptodata/2/iaa/6B.COMB.pep.*

5: /cgm2_6/ptodata/2/iaa/PCPUS.COMB.pep.*

6: /cgm2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225	8.1	582	4	US-09-091-725-17
2	217.5	7.9	494	3	US-08-660-645A-5
3	217.5	7.9	494	3	US-09-298-718-5
4	217.5	7.9	494	3	US-09-546-969-5
5	217.5	7.9	494	3	US-08-980-832-4
6	217.5	7.9	494	4	US-09-547-267-5
7	217.5	7.9	494	4	US-09-920-923B-4
8	200	7.2	414	3	US-09-067-626-4
9	191.5	6.9	492	1	US-07-783-705A-4
10	185	6.7	511	4	US-09-934-903-16
11	165.5	6.0	497	4	US-09-934-903-18
12	146	5.3	610	4	US-09-443-184-55
13	139	5.0	863	4	US-09-252-981A-19574
14	134	4.8	538	4	US-09-252-981A-30706
15	133	4.8	4928	3	US-09-036-987A-5
16	133	4.8	4928	3	US-09-370-700-5
17	133	4.8	4928	4	US-09-603-207-5
18	132	4.8	769	4	US-09-252-991A-17737
19	131.5	4.8	584	4	US-09-252-991A-28032
20	130	4.7	711	4	US-09-252-991A-22259
21	129.5	4.7	481	1	US-08-472-028A-6
22	129.5	4.7	481	2	US-08-808-323-6
23	129.5	4.7	481	3	US-09-050-603A-6
24	129.5	4.7	481	3	US-09-102-420B-6
25	129.5	4.7	481	3	US-09-015-683-6
26	129.5	4.7	481	3	US-09-497-698-6
27	129.5	4.7	481	4	US-09-497-698-6

28 129.5 4.7 483 3 US-09-071-296-6 Sequence 6, Appli
29 129.5 4.7 483 3 US-09-196-268-6 Sequence 6, Appli
30 129.5 4.7 483 4 US-09-191-938-6 Sequence 6, Appli
31 126 4.6 489 1 US-08-095-726-8 Sequence 8, Appli
32 126 4.6 489 1 US-08-095-726-10 Sequence 10, Appli
33 126 4.6 489 1 US-08-096-043-8 Sequence 8, Appli
34 126 4.6 489 1 US-08-096-043-10 Sequence 10, Appli
35 126 4.6 489 1 US-08-096-623A-8 Sequence 8, Appli
36 126 4.6 489 1 US-08-096-623A-10 Sequence 10, Appli
37 124.5 4.5 4302 4 US-09-052-469-8 Sequence 8, Appli
38 124.5 4.5 4302 4 US-08-422-582-8 Sequence 8, Appli
39 124.5 4.5 4302 4 US-09-052-262-8 Sequence 8, Appli
40 124.5 4.5 4339 4 US-09-052-469-6 Sequence 6, Appli
41 124.5 4.5 4339 4 US-08-422-582-6 Sequence 6, Appli
42 124.5 4.5 4339 4 US-09-052-262-6 Sequence 6, Appli
43 123.5 4.5 3562 4 US-09-679-279-14 Sequence 14, Appli
44 122.5 4.4 528 2 US-08-808-931-10 Sequence 10, Appli
45 122.5 4.4 528 3 US-08-808-323-10 Sequence 10, Appli

ALIGNMENTS

RESULT 1

US-09-091-725-17
; Sequence 17, Application US/09091725
; Patent No. 6329141
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Improved methods for transforming Phaffia
; and recombinant DNA for use therein
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster LLP
; STREET: 2000 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: United States of America
; ZIP: 20006-1898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,725
; FILING DATE: 23-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95203620.0
; FILING DATE: 22-DEC-1995
; APPLICATION NUMBER: EP 96200943.7
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: E. Victor Donahue
; REGISTRATION NUMBER: 35,492
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 582 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-091-725-17

Query Match 8.1%; Score 225; DB 4; Length 582;

Best Local Similarity 24.6%; Pred. No. 2e-11;

Matches 137; Conservative 65; Mismatches 228; Indels 126; Gaps 24;

QY 7 ATVCGSHALVSAATLAREGVSVEVELEKDTVLGAVSTVERFPQYKVDRCSSAHLM-IR 65

DB 12 AIVCGGIGGATLAREGQVTFVFEKNYSGRCSLIER-DGYRFDQGPSLILLPDL 70

QY 66 HSGIIEELGLGHHG-LRYIDCDP---WAFAPAPFGTDPGVFHRDLDAQCQSIERACGT 121

101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

RESULT 2

IS-08-660-645A-5

Sequence 5, Application US/08660645A

Patent No. 6087152

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/660,645A

FILING DATE: 07-JUN-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 9510888.9

FILING DATE: 09-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 494 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-660-645A-5

Query Match 7.9%; Score 217.5; DB 3; Length 494;

Best Local Similarity 23.6%; Pred. No. 7.3e-11;

Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

QY 7 AVVGSCHNALVSAAYLAREGWSVLEKDTVLGGAVSTVERPPGYKYVDRGSSAHLMIKH 66

DB 4 AIVIGAGFGGLALAIRLQSGAGIATTIVEARDKPGGR-AVVMNDQGHVFDAGPT---VVVD 59

QY 67 SGIIIEGLIGAGHLYIDCDPWAF-----PPAP-----GTDGFGIVFHRDLDA 110

DB 60 PDSUREL-----HALSQPMERDVTLTPVSFFYRLTWADGSRSEFYVNDDE 105

QY 111 TQOSIERACGTDADAYRRFVAVMSERH-VMAFSTP--PTGS--NLIGAPGGIATAR 165

DB 106 LIRQV-ASFPNADVDGTRRPHDYAEVYREGVYKLTCTTFLKLGOMLNAAPALMLRQAYR 164

QY 166 GNSLSRQFLAPGDALLDEYFDSEALKAAALWFGAGSGPMPSEPGTAPWVGFPAALMHVLP 225

DB 165 SVTSMVARFIQ--DPHLROAFSPHTLLV-----CGNPFST-----SIYALHALE 208

QY 226 PGRV-----GGSGALSAALASMAVDGATVALDGVTSIRRNNSHWT-VTTESGREVHAR 280

DB 209 RRGVWPAKGTNQLVAGMVALFERLGGTLLNARVTRIDTGDRAFGVTLDDGRLRAD 268

QY 281 KYIAGCHILTL-DILONGGDFRT-----TLDRRRKIRVGPICGAVLRLATSLPSYRG 334

DB 269 TVASNGDVHSHYRDLGHTTRGTGTAAILNRQWSMSL-----FVLHGLSKRPE--- 318

QY 335 DATTRETSGLQLVSDRAHLRTAHGAALAGELPPRAVLGMSFGS----- 380

DB 319 -----NLAHSHVIFG--PRYKGLVNEIFNGPRLPDDPSMYLHSP 355

QY 381 --IDPTIAPAG--RHQVTLWSQHPYRLSGHRDWSVAEAEADRVIGENAEFA-PGFTDS 435

DB 356 CVTDPSLADEGMSTHYVL---APVHLGRADVDEAEAFGYAERIPFELERRAIPDLRKH 412

QY 436 VLDRFIOTREDIESELGMIGGNVHMVMSLDQMLWRPLPELSGHRVPCADGLYLGTAST 495

DB 413 LTVSRIFSPADFTSELGAHGSASFVEPILTQSAWFRP-----HNRDRAIPNFYIVGACT 467

QY 496 HPGGVSG--ASGRSAARIALSD 516

DB 468 HPGAGIPGVVGSAAKATAQVWLSD 490

RESULT 3

US-09-298-718-5

Sequence 5, Application US/09298718

Patent No. 6124113

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

335 DATTRETSGLQLVSDRAHLRTAHGAALAGELPPRAVLGMSFSG----- 380
319 -----NLAHSVIFG--PRYKGLVNEIFNGPRLPDDFMSYMLHSP 355
381 --IDPTIAPAG--RHQVTLWSQMPYRLSGHRDWSVAEADRIVGNEMAFAPGFTDS 435
356 CVTDPSLAPGEMSTHYVL---APVPHLGRADVWEAEAPGYAEIRFEELERRAIPDLAKH 412
436 VLDRFCTQTRDIESELGMIGGNVHMVMSLDQMLWRPLPELSGHRVPGADGLYLTGAST 495
413 LTWSRIFSPADFTSELSAHGSAFSEVPILTQSAWFRP-----HNRDRAIPNFYIVGACT 467
496 HPGGVSAG--ASGRSAARIALSD 516
468 HPGAGIPGVVGSAAKATAQVWMLSD 490

RESULT 5
S-08-980-832-4
Sequence 4, Application US/08980832B
Patent No. 6291204
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/08/980,832B
CURRENT FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 494
TYPE: PR1
ORGANISM: Flavobacterium sp. R1534
S-08-980-832-4

Query Match 7.9%; Score 217.5; DB 3; Length 494;
Best Local Similarity 23.6%; Pred. No. 7.3e-11;
Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

7 AVVVGSHNALVSAAYLAEGMSVLEKDTVLGGAVSTVERPPGYKVDGSSAHLMTIRH 66
4 AVVIGAGFGLALALRLQAGIATTIVEARDKPGGR-AVVMNDQGHVFDAGPT---VVTD 59
67 SGIIELGLGAGHLRYIDCDPWAF-----PPAP-----GTGPGIVFHRDLDA 110
60 PDSIREL-----NALSQPMERDVTLLPVSPFRLTWADGRSFYVNDDE 105
111 TCOSIERAGCTKADAYRFBVAVMSERSH-VMKAFSTP--PTGS--NLIGAFGLATAR 165
106 LIRQV-ASFNPADVGYRFRHDYAEVREGYKLGTTPTFLKQMLNAPALMELQAYR 164
166 GNSLSRQFLAPGDALLDEYFDSALKALANFAGSQSGPPMSEPGTAPWGFALMHVLP 225
165 SVHSMVARIQ--DPLHROAFSFHTLLV-----CGNPFSTS-----SIYALIHAE 208
226 PGRAV---GGGSAALASAMVADGATVALGDGVTISIRRNSEWT-VTTESGREVHAR 280
209 RRGGVWFAGKGTNLVAGVVALPRLGGTILLNARVTRIDTSGDRATGVLLDGRQLRAD 268
281 KVTAGCHILTL-DLLNGGFPRT-----TLHWRKRVGPGICAVLRLATSLPSTRG 334
269 TVASNGDVMSHYRDLGLHTRGRKTRKAAILNRQWSML-----FVLHPLGSKRAPE--- 318
335 DATTRETSGLQLVSDRAHLRTAHGAALAGELPPRAVLGMSFSG----- 380
319 -----NLAHSVIFG--PRYKGLVNEIFNGPRLPDDFMSYMLHSP 355
381 --IDPTIAPAG--RHQVTLWSQMPYRLSGHRDWSVAEADRIVGNEMAFAPGFTDS 435
356 CVTDPSLAPGEMSTHYVL---APVPHLGRADVWEAEAPGYAEIRFEELERRAIPDLAKH 412
436 VLDRFCTQTRDIESELGMIGGNVHMVMSLDQMLWRPLPELSGHRVPGADGLYLTGAST 495

Db 413 LTWSRIFSPADFTSELSAHGSAFSEVPILTQSAWFRP-----HNRDRAIPNFYIVGACT 467
Qy 496 HPGGVSAG--ASGRSAARIALSD 516
Db 468 HPGAGIPGVVGSAAKATAQVWMLSD 490

RESULT 6
US-09-547-267-5
Sequence 5, Application US/09547267
Patent No. 6613543
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-547-267-5

Query Match 7.9%; Score 217.5; DB 4; Length 494;
Best Local Similarity 23.6%; Pred. No. 7.3e-11;
Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

7 AVVVGSHNALVSAAYLAEGMSVLEKDTVLGGAVSTVERPPGYKVDGSSAHLMTIRH 66
4 AVVIGAGFGLALALRLQAGIATTIVEARDKPGGR-AVVMNDQGHVFDAGPT---VVTD 59
67 SGIIELGLGAGHLRYIDCDPWAF-----PPAP-----GTGPGIVFHRDLDA 110
60 PDSIREL-----NALSQPMERDVTLLPVSPFRLTWADGRSFYVNDDE 105
111 TCOSIERAGCTKADAYRFBVAVMSERSH-VMKAFSTP--PTGS--NLIGAFGLATAR 165
106 LIRQV-ASFNPADVGYRFRHDYAEVREGYKLGTTPTFLKQMLNAPALMELQAYR 164
166 GNSLSRQFLAPGDALLDEYFDSALKALANFAGSQSGPPMSEPGTAPWGFALMHVLP 225
165 SVHSMVARIQ--DPLHROAFSFHTLLV-----CGNPFSTS-----SIYALIHAE 208

226 PGRV-----GSGGSAALASRMVADGATVALGDGVTISRENSNHT-VTTESGREVHAR 280
209 RRGVWFAGGNTQNLVAGWALFERLGOTLLNARVTRIDTEGDRATGVTLLDGRQLRAD 268
281 KVIAGCHLITL-DLLGNCGFDR-----TLDHRRKIRVCPGIGAVLRATSAALPSYRG 334
269 TVASNGDVHSTRDLGHTRRGRKTAAILNQRMSL-----FVLFGLSKRPE--- 318
335 DATTRETSGLQLLVSDRAHLRTAHGAALAGELPPRPDAVLGMSFSG----- 380
319 -----NLAHHSVIFG--PRYKGLVNEIFNGPRLPDDPSMYLHSP 355
381 --IDPTIAPAG--RHQVTLWSQWOPYRLSGHRDWSVAABADRIVGEAFA-PGFTDS 435
356 CVTDPSLAFBGMSTHYVL---APVPHLGRADVWEAEPGVAERIIFELERRAIPDLRKH 412
436 VLDRIQTPRIESELGMIGCNVHVMESLDQWMLRPLPELSGHRVPGADGLYLTGAST 495
413 LTVSRIFSPADPSTELSAAHGSASFVEPILTQSAWFR-----HNRDRAIPNFYIVGAGT 467
496 HPGGVUSG--ASGRSAARIALSD 516
468 HPGAGIPGVVGSASAKATAQVWLSD 490

SULT 7
-09-920-923B-4
Sequence 4, Application US/09920923B
Patent No. 6677134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/125944)
CURRENT APPLICATION NUMBER: US/09/920.923B
PRIORITY FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIORITY FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent in version 3.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
-09-920-923B-4

Query Match 7.9%; Score 217.5; DB 4; Length 494;
Best Local Similarity 23.6%; Pred. No. 7.3e-11;
Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

7 AVVVGSGHNLVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGRGSSAHLMIHR 66
4 AIVIGAGFGGLALAIRLSAGIATTIVEARDKPGGR-AVVMNDQGHVFDAGPT---VVTD 59
67 SGITIEELGLNAGLRYIDCPWAPA-----PPAP-----GTDGPGIVFHRDLDA 110
60 PDSUREL-----WALSQGMERDVTLLPVSPFYRLTWADGRSPFYVNDDE 105
111 TCQSIERACGTKDADAYRFVAVMSERSH-VMKAFSTP--PTGS--NLIGAPGGLATAR 165
106 LIRQV-ASFNPADVGRYRHDYAEVYRGYKLGCTTPTPLKGLQMLNAPALMRLOAYR 164
166 GNSFLSQFLAPGALLDEYFDSEALKAAALWFGAGSGPPMSEPGTAPVMVGFALMHVLP 225
165 SVHSKVARFIQ--DPHLRQAFSFHTLLV-----GGNPFSTS-----SYVALIHALE 208
226 PGRV-----GSGGSAALASRMVADGATVALGDGVTISRENSNHT-VTTESGREVHAR 280
209 RRGVWFAGGNTQNLVAGWALFERLGOTLLNARVTRIDTEGDRATGVTLLDGRQLRAD 268
281 KVIAGCHLITL-DLLGNCGFDR-----TLDHRRKIRVCPGIGAVLRATSAALPSYRG 334

Db 269 TVASNGDVHSTRDLGHTRRGRKTAAILNQRMSL-----FVLFGLSKRPE--- 318
QY 335 DATTRETSGLQLLVSDRAHLRTAHGAALAGELPPRPDAVLGMSFSG----- 380
Db 319 -----NLAHHSVIFG--PRYKGLVNEIFNGPRLPDDPSMYLHSP 355
QY 381 --IDPTIAPAG--RHQVTLWSQWOPYRLSGHRDWSVAABADRIVGEAFA-PGFTDS 435
Db 356 CVTDPSLAFBGMSTHYVL---APVPHLGRADVWEAEPGVAERIIFELERRAIPDLRKH 412
QY 436 VLDRIQTPRIESELGMIGCNVHVMESLDQWMLRPLPELSGHRVPGADGLYLTGAST 495
Db 413 LTVSRIFSPADPSTELSAAHGSASFVEPILTQSAWFR-----HNRDRAIPNFYIVGAGT 467
QY 496 HPGGVUSG--ASGRSAARIALSD 516
Db 468 HPGAGIPGVVGSASAKATAQVWLSD 490

RESULT 8
US-09-067-626-4
Sequence 4, Application US/09067626
Patent No. 6177086
GENERAL INFORMATION:
APPLICANT: Riley, Lee W.
APPLICANT: Nathan, Carl F.
APPLICANT: Ebrt, Sabine
TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/067,626
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/045,688
FILING DATE: 06-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/491
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-067-626-4

Query Match 7.2%; Score 200; DB 3; Length 414;
Best Local Similarity 24.5%; Pred. No. 2e-09;
Matches 113; Conservative 57; Mismatches 205; Indels 86; Gaps 19;

QY 72 ELGLGAGLRYIDCPWAPAPAGTGTGPGIVFHRDLDTCCSIERACGTKDADATRFV 131
Db 1 EFDLPARGVT-LTVPTIAYANPLFGR--PAAIAYHDLAHTCAKLD-----DGASWRRL 51

Y 132 AYWSESRHYWKAFTPTGSGNLGAFGLATARGN-SELSEQFLAPGD-----ALLDEYF 186
 b 52 GFLVAHSETWVEF-----MLSDKRSLEPTALGSLVRLGLRLMAQOTPAWRSLAGE-- 100
 Y 187 DGEALKAALAWFCAQSGPPMSBPPTAPMVGFPAALMHLVLPGRAGVGGALSAALASRMAY 246
 b 101 DAFALPTGVAHAHAISLPSLVAGAGLML--ATLAHSVGMPIPVGCTQAIADALIADLRA 158
 Y 247 DGATVALGCVTSIRNSNHWTTTSGREVIARVIAGCHILITLIDLLGNGGDFRTILD 306
 b 159 HGRGLAAGEVEITEPQESVVVFTAPTALLRVY-----RDKLP 195
 Y 307 H-----WRRKIRVGPFGICAVLRATLSAL-----PSYRGDATTRESTSGIQLLVSDRAHLRTA 358
 b 196 HRVAKALRYRFRAGIAKVDFVLSDEIPMSDFLRRRAATLH-----LGSTRDQWABA 247
 Y 359 HGAALAGSLPPRAVLGMSFGIDPTIAPAGHQVTLWSQWQPYRLSGHRDWSVAENAEA 418
 b 248 EADVAAGREHADPMVLAAC-----PHVADPGRIDET---GRAPFWTYAHVFGSGTDLA-T 298
 Y 419 DRIVGEMEAFAFGFTOSVL-DRFIQTPRDIESLGMIGNVMEVMSLDQMLMR-----P 473
 b 299 ETVTSVLERFAPGFRDIVAGRAVPAARMADENANYVGGDI-----TVGANSTWEAIAGP 353
 Y 474 LPELSGHRVPGADGLYLTCASTHPPGGVSGASGRSAARIAL 514
 b 354 TPLRLNWRTP-IPKVYLCSAATPPGAGVHGMCQWYAARTLL 393

RESULT 9

S-07-783-705A-4

Sequence 4, Application US/07783705A

Patent No. 5429939

GENERAL INFORMATION:

APPLICANT: Misawa, No. 5429939ihiko

APPLICANT: Kobayashi, Kazuo

APPLICANT: Nakamura, Katsumi

APPLICANT: Yamano, Shigeyuki

TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE

SYNTHESIS OF CAROTENOIDS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ladas & Parry

STREET: 26 West 61 Street

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10023

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: N/A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/783,705A

FILING DATE: 19911023

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1-103078

FILING DATE: 21-APR-1989

APPLICATION NUMBER: JP 2-53225

FILING DATE: 05-MAR-1990

APPLICATION NUMBER: US 07/519,011

FILING DATE: 19-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Schwadron, Janet I.

REGISTRATION NUMBER: 33,778

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-708-1935

TELEFAX: 212-246-5959

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

; LENGTH: 492 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-783-705A-4

Query Match 6.9%; Score 191.5; DB 1; Length 492;

Best Local Similarity 23.9%; Pred. No. 1.5e-08;

Matches 131; Conservative 65; Mismatches 242; Indels 111; Gaps 26;

Qy 9 VVSGHNALVSAAYLAREGMSVEVLEKDTVLGGAVSTVERPGYKVDGSSAHLIRHSG 68

Db 6 VIGAGFGGALAIRLQAAGIPVLLLEQRDKPGRAYVYED-CGFTFDAGPT---VITDPS 61

Qy 69 IIEEL-GLGAHLR-YIDCDP-----WAPAPAPGTDGPGIVPHRDLDAT-CQSIER 117

Db 62 AIEELPALAGKQKKEYVELLPVTPFVRLCWE-----SGKVFNVDNDQTRLEAQIQ 111

Qy 118 ACOTKADADAYRFVAVWSERSHYVK-----APSTPTGS--NLIGAFGLATARG-NSEL 170

Db 112 QFNPRDVEGYRQFL-----DYSRAVFKGYLKLGTGYFFLSFRDMLRAAPQLAKLQAWRSYV 167

Qy 171 SRQFLAPGDALLDEYFDSALKAALAWFG-AOSGPPMSEPPTAPMVGFPAALMHLVLPQGRA 229

Db 168 SK-----VASIIEDEHLRQAFSPHSLLVGNPFATS-----SIYTLIHALERENG 212

Qy 230 V-----GGSGALSAALASRMADVGCATVALGDGVTSTIRNSNH-WTVTTSGREVHARKVIA 284

Db 213 VMPFRGGTGAIVQGMIKLPQDLGGEVILNARVSHMETTGNKIEAVHLEDGRFRFLTAQAVAS 272

Qy 285 GCHILITL-DLLGNGGDFRTTLQHWRRKIRVGPFGICAVLRATLSALPSYRGDATTREBTS 343

Db 273 NADVVHTYRDDLSQHPAAVQSNKLTQX-----RMSNSLFVLYFG----- 312

Qy 344 GLQLLVSDRAHLTAHGAALAGELPPPAVL-----GMS--FS-----GIDPTIA 386

Db 313 -----LNHHHDOLAHHTVCFG---PRYRELIDEIFNHDGLAEDFSLYHAPCVTDSSIA 363

Qy 387 PAGHQVTLWSQWQPYRLSGHRDWSVAEAEADRIVGEAE-PAFGTDSVLDRFIQTPR 445

Db 364 PEGCGSYVVLAP-VPHLGTANLDWTVEGPKLDRIFAYLEQHYMFGLRSQLWTHRMFTPF 422

Qy 446 DIESELGMIGNVMHVMESLDQMLMRPLFELSGHRVPGADGLYLTCASTHPPGGVSGAS 505

Db 423 DFRDQLNAYHGSASFVEPVLTSAPWRP-----HNRDKTITNLVLVGAGTHGAGIPGVI 477

Qy 506 GRSARIAL 514

Db 478 GSAKATAGL 486

RESULT 10

US-09-934-903-16

; Sequence 16, Application US/09934903

; Patent No. 6660507

; GENERAL INFORMATION:

; APPLICANT: Koffas, Mattheos

; APPLICANT: Odom, J. Martin

; APPLICANT: Schenzle, Andreas J.

; APPLICANT: No. 6660507ton, Kelley C.

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Rouviere, Pierre

; APPLICANT: Picataggio, Stephen

; APPLICANT: Cheng, Qiong

; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

; FILE REFERENCE: C11646 US NA

; CURRENT APPLICATION NUMBER: US/09/934,903

; CURRENT FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/229,907

; PRIOR FILING DATE: September 1, 2001

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 16

```

; SEQ ID NO 18
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ORF9
US-09-934-903-18

Query Match      6.0%; Score 165.5; DB 4; Length 497;
Best Local Similarity 23.3%; Pred. No. 3e-06;
Matches 128; Conservative 75; Mismatches 210; Indels 137; Gaps 30;

Qy      8 VVVGSHNALVSAAYLARGSGWSEVLEKDTLGGAVSTVERPPGKYKVDGSSAHLMIHRS 67
Db      10 IVIGAGLGLSALISATAGFSVOLIEKNDKNGKLNIMTK-DGFTFLGPSI-LTMPH- 66
Qy      68 GITEEL--GLGAGLGLYVID---CDP-WA-FAPPARGTD-----CPGIV-- 103
Db      57 -IPEALFTGAGKNWADYVQIQVEPHRNFPEDGSGVIDLCEDAETQREKDLKLGPIYVQ 128
Qy      104 FHRDLDT--COSIERAGCTKDAAYRRFVAVWGSERHVNKAFPTPTPGSNLIGAPGG 160
Db      126 FQRFIDYSKNLCTETETAGYFAKELDGF-----W-----DILKTYGP 161
Qy      161 LATARGNSLSRQFLAPGDALLDEYFDSEALKAAALWFGAQSPPMSBPGTAPMVGPAAL 220
Db      162 L-----RSLLSFDVFSMDQGVRRFLISDEKLVEILNYIKYVG---SSPYDAP-----AL 208
Qy      221 MEVLPPGR-----AVGSGSALSAALSRMAYD-GATVALGDGVTSIRRNNSHWTVTTE 272
Db      209 MNLIPYIQVHYGLWYKGGMYGQAQAM-EKLAVELGVETRLDAEYSEIQK-----Q 258
Qy      273 SGREYHARKVIACHILITLIDLLGNGGDFRITLDHWRKKIRVPGICGAVLKLATSALPSY 332
Db      259 DGR-----ACAV-----KLANG--DVLPAIVSNMEVIFAMEKILRSASELKQM 302
Qy      333 RGDATTRESTSGHLQ-LVSDRAHLTAHGAALAGELPPR-----PAVLGMS 377
Db      303 Q---RPEPSCGLVHLGLGDRLYLPOLAHNNPFYSDHPREHFDVFKSHERLSDDPTIYLV 359
Qy      378 FSGIDPTIAPAGHQVTLW---SQWQPYLSHRDWSVAEAEADRIVGEAEAPGFTD 434
Db      360 PKTDPQAAGACGEEIKILPHIPHDPKLITAEYDYSALRE----RVLVKLRM--GLTD 413
Qy      435 ---SVLDRFTQTPDIESLSLGMIGGNVHVMSELDQMLMRPLPELSGHRVPGADGLYLT 491
Db      414 LRQHVITEEWTPLDIOAKYNSQSGIYGVAD-----RPNKILGFKAPQRSELSNLVYV 468
Qy      492 GASTHPGGGV 501
Db      469 GGSVNPGGGM 478

RESULT 12
US-09-443-184-55
; Sequence 55, Application US/09443184A
; Patent No. 6372431
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger Gary
; APPLICANT: Kaseer, Matthew R.
; APPLICANT: Panzer, Scott
; APPLICANT: Seilhammer, Jeffrey J.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah
; APPLICANT: Azimzai, Mariad
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
; FILE REFERENCE: PC-0007 US
; CURRENT APPLICATION NUMBER: US/09/443,184A
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program

```

RESULT 12
US-09-443-184-55
; Sequence 55, Application US/09443184A
; Patent No. 6372431
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Panzer, Scott
; APPLICANT: Seilhammer, Jeffrey J.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah
; APPLICANT: Azimzai, Valda
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
; FILE REFERENCE: PC-0007 US
; CURRENT APPLICATION NUMBER: US/09/443,184A
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program

SEQ ID NO 55

LENGTH: 610

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No. 6372431 1867333CD1

IS-09-443-184-55

Query Match 5.3%; Score 146; DB 4; Length 610;
Best Local Similarity 23.6%; Pred. No. 0.00021;
Matches 134; Conservative 73; Mismatches 239; Indels 122; Gaps 29;

5 LDVVVGSCHNALVSAAYLAREGWSVEVELEKOTVIGGAVSTVERPFGYKVDGSSAHLMI 64
68 LDVVVGSFGGLAAALAKAGKRVLVLEQHTKAGGCCCTFGK-NGLEFFDTG-----I 120
65 RHSGIIBELGLGAGHLYIDCDWAPAPPAGTD-----GPG-----IVHRDLDTACOSI 115
121 HYIGRMEEGIGRFLDQITQEGGLDWAFLSSPDDIMVLESGNGRKEYPMYSEKAYIQGL 180
116 ERACGTKDA--DAYRFVAVMSRSHVMKAFSTPTGNSNLIGAFGLATARGNELSRQ 173
181 KEKFPQOEAIIDKYIKLVKVVSSGAPHAIIKFLPLPVVQLLDRCGLLTFRFSFPLQASTQ 240
174 FLAPGDALDEYFDSALKALAWFGAQS--PPMS-----EPGTAPMVGPA 218
241 SLA---EVLQGLGASSELQAVLSYIPTVGTNHSAPSHWALLVNHVMKGGFYERGSS 297
219 AL-MHVLPPGKAVGSGSALASMAVDAVATVAGDGVTSIRNSNHMTVTTSGRE- 276
298 EIAFHTPIVIOAGGA-VLTKATVQSVLLDSAGKAG-----VSVKKGHEL 342
277 --VEARKVIAGCHILTLDLLNGGDFRTTLDHWRKI-RVGGIGAVLRATLSALPSVR 333
343 VNIYCIPIVSNAGLFNTYHLLPG--NARCLPGVKQOGLGTVRPGLG-----MTSFIQLR 395
334 GDATRE-----STSGIQLLVSRRAHLRTAHGAALAGELPPRAVLGMSF-SGIDPTIAP 387
396 G---TKEDLHLPSTNYVYVYDMDQWERYVSMPEEAHEIPLLFAPPSAKOPTWED 452
388 --AGRHQVTL-----WSQWQPYRLSGHR--DWA-----SVAREAD-----RIVCE 424
453 RFGKSTMIIMLPTAYEWPEMOA-ELKGRGSDYETFXNSFVEASMSVVLKLFQLECK 511
425 MEAFAPGFTSVLDR--FIQTPR-----DIESELGMIGNNVHVMESLDQMLWRPLPELS 478
512 VESVTAQ---SPLTNQPYLAAPRGACYGADHDLDRLHPCVM-----ASLRQAQSPIP--- 559
479 GHRVPGADGLYLTGASTHPGGVSGASG 506
560 -----NLVLTQDIFTGCLVCAQQ 579

RESULT 13

US-09-252-991A-19574

Sequence 19574, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19574

LENGTH: 863

TYPE: PRT

Query Match 4.8%; Score 134; DB 4; Length 538;
Best Local Similarity 20.9%; Pred. No. 0.002;
Matches 123; Conservative 73; Mismatches 187; Indels 206; Gaps 30;

5 LDVVVGSCHNALVSAAYLAREGWSVEVELEKDTVLGCAVSTVERPFGYKVDGSSAHLMI 64
40 VDMLLVAG---IMSATL-----AVLLKELDPNLKWE-----VVEL 72

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19574

Query Match 5.0%; Score 139; DB 4; Length 863;

Best Local Similarity 22.4%; Pred. No. 0.0015;

Matches 114; Conservative 51; Mismatches 187; Indels 156; Gaps 25;

87 PWAFAPPAGCT--DQPGIVPHRD-----LDATQOSIRACGTKDADAYRFFVAVWSERSH 140
36 PRFSPSTSANPDGPMVSNTPSPISPPSAGPIRSGSSTSTS-----AIIATWT----- 86
141 VMKAFSTPTGNSNLIGAFG-GLATARGNELSRQFLAPGDALLDEYFDSALKAL--A 196
87 -----TGT-----GANPCGLECSAGSSCRNRRWHERPA-----ARRLGAGLRAP 127
197 WPGAQSGPPMSEPGTAPMVGFAALMHVLPGRVAGSGSALASMAVDAVATVAGD 256
128 WLSARVRRREADPGVPFAVAF-----PGRHAGAAG--PGAVGRNRNSTGIRRLGTR 178
257 VTSIRENSNHMTVTTSGREHV-----ARKVIAGCHILTLDLLNGGDFR 302
179 RTAVRPTLR---CRTSVRPVHLPLWLSGRADPPYRPGHQRGACPAITGRP--GQGRGER 233
303 TTLDMRRKIRVOPGICAVLRATLSALPSYRGDATTRESTGQLLVSDRAHLRTAHG-- 360
234 -----RPVVRTPGAGPRLRRPAGGVPRLYGA-----CAARRAVRPGPLP 275
361 AALAGELPPRPVAVLGHSFSGIDPTIA-----PAGRHQVTLWSQWQPYRLSGHRDWASV 413
276 APTAARLPAER--LAARYAGRRRTVAASAGRRHPAGRHRL----- 313
414 ABAEADRIVE-----MEAFAPGFTSVLD-----RFIOTPRDIESELGMIG 455
314 -RAAAGGLAGTGGWLDLISGVDCRVATSPGREDPRLEDVHRGRVPPPAHDQRMAGPG 372
456 GNNVHVMESLDQMLWRPL-----PELSGHRV-----PGADGLYLTGASTHPGGV 501
373 GRPGVSAYAVARL--RPAHRAQRPDRAARMACGGRQPCRPDPHVPVLRQAGSSHPGTGR 430
502 SGAS-----GRSAARIALSDSRGKAS 523
431 GGAAAVPVPPGAAAGEHGLVADRFQGPS 458

RESULT 14

US-09-252-991A-30706

Sequence 30706, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 30706

LENGTH: 538

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30706

65 RHSGIIEELGLGHGHLRYIDCDPWAPFAPPACTGGPGIVFHRRDLDTATCCSIERACGTDA 124
:
:
:
73 QESGAIE-----SSNPMKNA-----GTGHAGLC---ELNYTPQSDGSIDIKA 113
:
:
:
125 DAVRRPVAYMSRSHVMK--AFSTPTGSLIGAFGLLATARGNELSRQLPAGDALL 182
:
:
:
114 VGINTWFEVSKQFWSHLVAKTFCSPKTFINVPFLHSVRGSEGIYALKRPES-----L 168
:
:
:
183 DEYPDSEAL-----KAALA-WFGAOSGPPMSEPOTAPMWGFPAALMEVLPPGRAVGSS--- 233
:
:
:
169 TKHIAFETWYGEDKTAEWM-----PLMPGRPADEAIAT-----RVEGTDVN 215
:
:
:
234 -GALSAAASRAVMD-GATVALGDGVTSIRNSNHWTVT---TESG-REVHARKVIACC 286
:
:
:
216 FGALTUOLLOHLAQEQCAIQRYNQKTHLRADNGRWTVKDTRANGCDREIQARFV---- 271
:
:
:
287 HILUTLLDGNGGFORTTLDHWRRIKVGPDIGAVRLATSALPSVYRGDATRESTSGLO 346
:
:
:
272 ---FLCAGG-----GALPLLQLSGPIPEGKG----- 293
:
:
:
347 LLVYSDBLAHTAHGAALAGELSPRAVLGMFSFGIDPTIAPAGRHOVTLMOWO---EVR 403
:
:
:
294 -----FCGFP-----VSGWRCNDNEIVK--CHQAKVYQRHVCSPPM 330
:
:
:
404 LSGHRDWASVAEAEDRIVGEMEFAPGFTDSVDLFRTQTFRDIETSELGNIGNNM-HVE 462
:
:
:
331 SVPHLD-TRVWDKKSLLPYPAGFS-----TKFLRHGHSFLDLPLSVRPCNILPMLS 381
:
:
:
463 MSLDQMMLWRPL-----PELSCHRNV-----PCADOL 488
:
:
:
382 VARONMDLTRYLIGQYMOSPQRLEALRKFYPEARABDRLEVAGQVRVQIKDKPKKGGI 441
:
:
:
489 YLTG---ASTHPG-----GGVSGSAGRSAAIALSD---SRRGKSQW 525
:
:
:
442 LQFGTELVAARDQSIBALLGASFGSALVTVISIMLGIERCPQEARSPEW 490
:
:
:

RESULT 15

3-09-036-987A-5
Sequence 5, Application US/09036987A
Patent No. 6143526

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive

TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide

TITLE OF INVENTION: Production

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dow AgroSciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479

```

; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4928 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-036-987A-5

Query Match 4.8%; Score 133; DB 3; Length 4928;
Best Local Similarity 23.4%; Pred. No. 0.062;
Matches 151; Conservative 61; Mismatches 196; Indels 238; Gaps 39;

QY 3 AFLD--AVVVGSGHVALVSAAYLAEGNSV-EVLEKDTVLGGAVSTV-----ERPFG--- 51
DB 531 AALDERAVVLGSDRAALLTGLRAFADGCDPAPEVVGSGVLGGRVGVFSGQGQWPEGMR 590
QY 52 --YKV-----DR-----GSSAHLMIR-----HSGIIE-ELG-- 74
DB 591 GLYSVFPVPADAFDEACAEALDHLGOELNRVDFVGSQAWLLDRTVWAGSGLFAQLIGLL 650
QY 75 --LGAGHLRYIDCDPWAPAPAGPTDGPICVFHRDLDATCSIERACGFKDADAYRRFA 132
DB 651 RLLGSGWR-----PDVVLGHSVGEL--AAVHAAGVLSLSEAAALVA 690
QY 133 VMSESRHWKAPSTPTGSLNLIAGPGLATARGNSLSRQFLAPODALLIDYDFDSEALK 132
DB 691 ---GRAR-LMQAL--PSG---GAM--LAVATGFEQV-----DPLLD----- 720
QY 193 AALAWRGAQSGPMSEPGTAPVMVGPAAALMHVLPPGRAVGGSGALSAAALSRMAVDGATVA 252
DB 721 -----GYRDRIGIAVNG--PESVVLSDRELLTEIADRLHDQCR-- 759
QY 253 LGDGVTSIR-----NSNHWVTYES-----GREVHARKVIAGCHILITLILGNGGPD 301
DB 760 -----TRWLSVSHAFESPHEPMLKEFAQISRGREYHAPEL-----PIISTLIGELDGRV 810
QY 302 RTTLDHWRKIR-----VGPFGVAVLRATLSALPSVGRGATTRFS--TSLQL 347
DB 811 MGTPEVWVQVREPVREFAEGVQALVQGVGTIVELGPDGALSTLVEECVAESGRVAGIPL 870
QY 348 LVSDRAHLRTAHGAALA-----GELPPRAVLGMSFGSID-PTIAPAGRHQVTLWSQW 400
DB 871 MKRORDAEATVL--AALAAQIHTRGGEVDWRSFFAGTGAKQVDLPYA-----FORQ 919
QY 401 PYRLSG---HRDWASVAREABDR-IVGENEAFAPGPTSDVLDRFIOTPRDIESELMIGG 456
DB 920 RYWLASTGRAGDVTAAGLAEADHPILGAVVALADG-----EGVLTGR-----LTAG 966
QY 457 NUNHIVEMSLDQMLWPLPPLSELSHRVPG---ADGLYLTGASTHSG----- 498
DB 967 S---H-----FWLSDRHVGEIVVPGTAIVELVHWGRLGGRVEELALEA 1010
QY 499 -----GGVS-----GASGRSAARIALSDSRGKA--SQWMRRSS 530
DB 1011 PLILPDHGAVOVVLVGPPECSGARSVALYSCPGEAIEPEWKKHAT 1056

```

Search completed: February 29, 2004, 14:55:08
Job time : 19.1665 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

! protein - protein search, using sw model

in on: February 29, 2004, 14:26:38 ; Search time 62.8999 Seconds
(without alignments)
2389.754 Million cell updates/sec

file: US-09-941-947a-38

irect score: 2768

quence: 1 MSAPLDAVVGSGHNLVSA.....ALSDSRGKASQAMRRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

al number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : A Geneseq_29Jan04.*

1: geneseq1980s.*

2: geneseq1990s.*

3: geneseq2000s.*

4: geneseq2001s.*

5: geneseq2002s.*

6: geneseq2003as.*

7: geneseq2003bs.*

8: geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2768	100.0	532	5	Aae22317 Rhodococc
2	2764	99.9	532	6	Abp71892 R. erythr
3	859.5	31.1	511	6	Abp71893 D. radiod
4	688	24.9	542	6	Abp71894 Synchpcy
5	624.5	22.6	581	4	Am93657 Human pol
6	624.5	22.6	581	5	Abb7177 Human oxi
7	403	14.6	504	7	Rdc31532 Human nov
8	305.5	11.0	322	7	Rdc3152 Human nov
9	296.5	10.7	471	4	Asg90287 C glutami
10	270	9.8	501	6	Asc22615 Synchocy
11	225	8.1	582	2	Asw22499 Phaffia d
12	225	8.1	615	6	Asc22614 Tomato pl
13	221.5	8.0	548	4	Abb85728 Enzyme in
14	219.5	7.9	548	4	Abb76640 Coryneb
15	219.5	7.9	548	4	Abb76641 Coryneb
16	219.5	7.9	548	4	Asg90440 C glutami
17	217.5	7.9	494	2	Asw00871 Flavobact
18	214.5	7.7	494	2	Asw69532 Flavobact
19	200	7.2	411	2	Asw83358 Mycobacte
20	198.5	7.2	494	6	Am70123 Photorhab
21	195.5	7.1	544	5	AAO15518 Agromyces
22	192.5	7.0	530	6	Asc31690 Rhodococc
23	191.5	6.9	492	2	Aar07466 Polypepti
24	191.5	6.9	492	2	Asw82257 C. utilis
25	191.5	6.9	492	2	Asw87889 Protein e

26	191.5	6.9	492	2	AAW99099	Erwinia u
27	191.5	6.9	492	2	AAW26333	Erwinia u
28	189.5	6.8	492	5	AAE22314	Pantoea s
29	189.5	6.8	492	5	AAO16021	Pantoea s
30	189.5	6.8	492	6	ABP96688	Pantoea s
31	185	6.7	511	5	AAE22309	Methylo
32	185	6.7	511	5	ABG61588	High grow
33	185	6.7	511	5	AAU80332	Methylo
34	185	6.7	511	5	ADA14534	Methylo
35	182.5	6.6	502	6	ABU43877	Protein e
36	182	6.6	121	4	AAU86860	Novel hum
37	182	6.6	121	7	ADB60194	Connectiv
38	182	6.6	595	5	ABG90884	Herbicida
39	181	6.5	587	6	AAO22616	Arabidops
40	173.5	6.3	205	4	AAW95800	Human rep
41	173.5	6.3	205	4	ABG96331	Human tes
42	167.5	6.1	497	6	ABU16031	Protein e
43	167.5	6.1	526	2	AAW95697	Erythroba
44	165.5	6.0	497	5	AAE22310	Methylo
45	165.5	6.0	497	5	ABG61589	High grow

ALIGNMENTS

RESULT 1
AAE22317
ID AAE22317 standard; protein; 532 AA.

XX

AC AAE22317;

XX

DT 25-JUL-2002 (first entry)

XX

DE Rhodococcus erythropolis AN12 beta-carotene ketolase (CrtO) enzyme.

XX

XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet; anti-oxidant; steroid; flavour; fragrance; electro-optic application; aquaculture; enzyme; beta-carotene ketolase; CrtO.

XX

OS Rhodococcus erythropolis.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "Encoded by GTG"

XX

PN WO200218617-A2.

XX

PD 07-MAR-2002.

XX

PF 04-SEP-2001; 2001WO-US027420.

XX

PR 01-SEP-2000; 2000US-0229858P.

PR

PR 01-SEP-2000; 2000US-0229907P.

XX

PA (DUPO) DU PONT DE NEMOURS & CO B I.

XX

PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

PI Odom JM, Picataggio SK, Rouviere PE;

XX

DR WPI; 2002-351711/38.

DR

DR N-PSDB; AAD35515.

XX

PT Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates.

PT

PT Claim 27; Page 148-150; 156pp; English.

XX

PS The invention relates to a method for producing carotenoid compounds. The method comprises a transformed metabolising host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the

CC

C control of regulatory sequences, and contracting the host cell with carbon
 C substrate to produce a carotenoid compound. The method is useful for
 C producing carotenoid compounds such as anthraxanthin and astraxanthin, by
 C using microorganism having a nucleic acid molecule encoding enzymes in
 C the carotenoid biosynthetic pathway and which metabolise single carbon
 C substrates. The carotenoids have potent anti-oxidant properties useful in
 C diet, and aquaculture elements. The carotenoids are also useful as
 C intermediates in the synthesis of steroids flavours and fragrances and
 C compounds for potential electro-optic applications. The present sequence
 C is Rhodococcus erythropolis AN12 beta-carotene ketolase (CrtO) enzyme
 C used in the invention

X Sequence 532 AA;

Query Match 100.0%; Score 2768; DB 5; Length 532;

Best Local Similarity 100.0%; Pred. No. 5.7e-223;

Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MSAPLDVAVVSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGSSA 60
 b 1 MSAPLDVAVVSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGSSA 60
 Y 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAPGTDGPGIVFHRDLDTATCQSIERACG 120
 b 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAPGTDGPGIVFHRDLDTATCQSIERACG 120
 Y 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180
 b 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180
 Y 181 LLDVDFDSEALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPFGRAVGGSGALSAAL 240
 b 181 LLDVDFDSEALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPFGRAVGGSGALSAAL 240
 Y 241 ASRMAVDGATVALGDGVTISIRNSNHHVTITTESGREVHARKVIAGCHILTLTDLGNGGF 300
 b 241 ASRMAVDGATVALGDGVTISIRNSNHHVTITTESGREVHARKVIAGCHILTLTDLGNGGF 300
 Y 301 DRTTLDHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360
 b 301 DRTTLDHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360
 Y 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420
 b 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420
 Y 421 IVGEMEAPAGTDSVLDRTFTQTRDIESBLGIMGVMHVENSLDQMLWRPLPELSGH 480
 b 421 IVGEMEAPAGTDSVLDRTFTQTRDIESBLGIMGVMHVENSLDQMLWRPLPELSGH 480
 Y 481 RVPAGDGLYLITGASTHFGGVSAGSRAARIALSDSRKASQWRRSSRS 532
 b 481 RVPAGDGLYLITGASTHFGGVSAGSRAARIALSDSRKASQWRRSSRS 532

RESULT 2

ABP71892

CD ABP71892 standard; protein; 532 AA.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PF 01-AUG-2002; 2002WO-US024317.
 XX 02-AUG-2001; 2001US-0309653P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Cheng Q, Tao L;
 XX WPI; 2003-300493/29.
 XX DR N-PSDB; ABZ75460.
 XX
 PT New isolated nucleic acid encoding carotenoid ketolase enzyme, useful for
 PT producing cyclic ketocarotenoid compounds such as adonirubin, echinenone,
 PT and as probes or primers to identify nucleic acids encoding the enzyme.
 XX
 PT Claim 4; Page 65-67; 90pp: English.
 XX
 CC The invention relates to a novel isolated nucleic acid encoding a
 CC carotenoid ketolase enzyme which: (i) encodes an amino acid sequence
 CC containing all six conserved motifs of CrtO enzymes of Rhodococcus
 CC erythropolis AN12 strain, Deinococcus radiodurans R1 strain, and
 CC Synechocystis sp. PCC6803 strain; (ii) encodes a sequence of 532 amino
 CC acids, given in specification; or (iii) hybridises to (i) or (ii). A
 CC nucleic acid of the invention is useful for obtaining a nucleic acid
 CC encoding a carotenoid ketolase enzyme. A nucleic acid of the invention is
 CC also useful for producing cyclic ketocarotenoid compounds. A chimeric
 CC gene is useful for regulating cyclic ketocarotenoid biosynthesis in an
 CC organism by introducing the gene into a host cell and growing the host
 CC cell under conditions whereby the carotenoid ketolase gene is expressed
 CC and the cyclic ketocarotenoid biosynthesis is regulated. The regulation
 CC may be upregulation of cyclic ketocarotenoid biosynthesis, where the
 CC carotenoid ketolase gene is overexpressed on a multicopy plasmid or is
 CC operably linked to a inducible or regulated promoter. Optionally the
 CC cyclic ketocarotenoid biosynthesis may be down regulated, where the
 CC carotenoid ketolase gene is expressed in antisense orientation or is
 CC disrupted by insertion of foreign DNA into the coding region. The present
 CC sequence represents the R. erythropolis strain AN12 crtO polypeptide
 XX

SQ Sequence 532 AA;

Query Match 99.9%; Score 2764; DB 6; Length 532;

Best Local Similarity 99.8%; Pred. No. 1.2e-222;

Matches 531; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSAPLDVAVVSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGSSA 60
 DB 1 VSAFLDVAVVGSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGSSA 60
 QY 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAPGTDGPGIVFHRDLDTATCQSIERACG 120
 DB 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAPGTDGPGIVFHRDLDTATCQSIERACG 120
 QY 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180
 DB 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180
 QY 181 LLDVDFDSEALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPFGRAVGGSGALSAAL 240
 DB 181 LLDVDFDSEALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPFGRAVGGSGALSAAL 240
 QY 241 ASRMAVDGATVALGDGVTISIRNSNHHVTITTESGREVHARKVIAGCHILTLTDLGNGGF 300
 DB 241 ASRMAVDGATVALGDGVTISIRNSNHHVTITTESGREVHARKVIAGCHILTLTDLGNGGF 300
 QY 301 DRTTLDHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360
 DB 301 DRTTLDHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360
 QY 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420
 DB 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420
 QY 421 IVGEMEAPAGTDSVLDRTFTQTRDIESBLGIMGVMHVENSLDQMLWRPLPELSGH 480

C *Synechocystis* sp. PCC6803 strain; (iii) encodes a sequence of 532 amino acids, given in specification; or (iii) hybridises to (i) or (ii). A nucleic acid of the invention is useful for obtaining a nucleic acid encoding a carotenoid ketolase enzyme. A nucleic acid of the invention is also useful for producing cyclic ketocarotenoid compounds. A chimeric gene is useful for regulating cyclic ketocarotenoid biosynthesis in an organism by introducing the gene into a host cell and growing the host cell under conditions whereby the carotenoid ketolase gene is expressed and the cyclic ketocarotenoid biosynthesis is regulated. The regulation may be upregulation of cyclic ketocarotenoid biosynthesis, where the carotenoid ketolase gene is overexpressed on a multicopy plasmid or is operably linked to a inducible or regulated promoter. Optionally the cyclic ketocarotenoid biosynthesis may be down regulated, where the carotenoid ketolase gene is expressed in antisense orientation or is disrupted by insertion of foreign DNA into the coding region. The present sequence represents the *Synechocystis* sp. strain PCC6803 crto polypeptide

X Q Sequence 542 AA;

Query Match 24.9%; Score 688; DB 6; Length 542;
Best Local Similarity 31.9%; Pred. No. 9.1e-49;
Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

Y 6 DAVVSGHNLVSAAYLAREGWSVEVLEKDTVLGGAVST-...VERPFGYKVDGSSAH 61
b 5 DAVIIGAGHGLVCAAYLLQRLGVLLEKREVPGGATTEALMPELSPQFRNCAIDH 64
Y 62 LMRHSGIIBELGHAHGLYIDCDPWAFAPAPGTDGPIVHRDLDTQOSIERACGT 121
b 65 EPIFLGVLQELMLAQVLEYLEFCDPSVT-...CPGLDGAQFMSYSLKTCAH-ATYSP 120
Y 122 KDADAYRRFVAVWSERSRHVMKAFSTPTGT-...SNLIGAFGLATARGNSBLS 171
b 121 RDAEKYQFVNYWTDLLNAVQFAPNAPQALLDLALNYGWNLSVLATAGSKTALDFI 180
Y 172 RQFLAGDALLDYFSEALKAAALWFGAQSGPMPSEPTCAPWGFALMHVLPGRVAVG 231
b 181 RTWIGSPEDVINEWFSERVKAPLABLCSEICAPPQSGKSSGMMVAMHLEGIARPKG 240
Y 232 GSGALSAALASMAVDGATVALGDGVTISR-RNSNHWITVTSRGREHARKVIAGCHILT 290
b 241 GTQALTEALVQLVQAQGGKLTDTQVTRVIVENNAIGVEVANGEQYRAKK-...GVIS 295
Y 291 TLD-...LLNGGDFRDTLTHWR-R-KIRVGPIGAVLRL-ATSALPSYRGDATTRE 340
b 296 NIDARREFLQVFGALAKVQNGLERLERTVNNNEALIKIDCALSGLPHTAMAGP-E 354
Y 341 STGCLQLLVSDRAHLRTAHGAALAGELPRPVLGNFSGIDPTIAPAGRHQVTLWQ- 398
b 355 DLRTILIAISVRHVEAHALIALGQIPDANPSLYLDIPTVLDPTWAPPGQH--TWIEF 412
Y 399 WQYRLSG-...HRDWA-SVAEAEADRIVGEMAFAPGTFDVLDFIOTPRDIESE 450
b 413 PAPYRIAGLGTLGNGWTDELKEKVAADRVIDKLDYAPNLKSLIIGRVESPAELAQ 472
Y 451 LGMIGNVHVSNDOMLWRPLPELSGHRVPGADGLVLTGASTHPGCVSGAGRSAA 510
b 473 LGSYNGNVYHLDMSLDQMFLEPPEIANQTF-INKYLYLTGATGTHPGGSGMGFRNCA 531
Y 511 RIALLSDRR 519
b 532 RVFLKQQR 540

RESULT 5

AM93657

ID AM93657 standard; protein; 581 AA.

CC AM93657;

CC AM93657;

CC 06-NOV-2001 (first entry)

CC Human polypeptide, SEQ ID NO: 3528.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
XX EP1130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000BP-00114089.
XX 08-JUL-1999; 99JP-00194486.
XX 11-JAN-2000; 2000JP-00118774.
XX 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
XX Ota T. Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX DR N-PSDB; AAK94592.
XX 830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.
XX Claim 8; SEQ ID NO 3528; 1380pp + Sequence Listing; English.
XX The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3'-ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO
XX Sequence 581 AA;

Query Match 22.6%; Score 624.5; DB 4; Length 581;

Best Local Similarity 33.0%; Pred. No. 2.1e-43;

Matches 185; Conservative 81; Mismatches 226; Indels 69; Gaps 18;

QY 6 DAVVSGHNLVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGSSAHLMIR 65
b 36 DAVVIGHGHLVAAAYLQRLGVTAVFERRHVGGAATELIIPGKFSRASYLLSLR 95
QY 66 HSGIIBELGHAHGLYIDCDPWAFAPAPGTDGPG-...IVFHRDLDTQOSIERAC 119
b 96 PQ-IYTDLELKHGLRLHRLNPYSFTPEL--BEGAGSKVPRCILLGTDMAENQKQIAQ-F 151
QY 120 GTKDADAYRRFVAVWSERSRHVMKARSTPTGT-...SNLIGAFGLATARG-... 166
b 152 SQDQAQVFPKTEEFMERLALADPLDAAAPVMAAPQHGSLLOMRSLSTLPLKXAGRI 211
QY 167 -NSELARQF-...LAPGDALIDYFSEALKAAALW-...FCAQSGPMPSEPTAPWGPAA 219
b 212 LGAQLPRYEVLTAPITKVLDQWFBSEPLKATLATDAVIGAMTSP--HTPGS-...GYVL 265
QY 220 LMRHVP-...GRVGGGALSAALASMAVDGATVALGDGVTISRNSNHW-...V 269
b 266 LHHVNGLBGMQAGWYVQGMGALSDAIASSATHTGASIFTEKTVAKQVNSEGCVQGV 325
QY 270 TTESGREVHARKVIAGCH-...ILTLDLLNGGDFRDTLTHWRKIRVGPVIGAVLR 323
b 326 VLEDGIEVRSKMWLSNTSPQITPLKTPQELPPEBFLERISQLDTRSPV-...TKIN 378
QY 324 LATSALPSY-...RGDATTRESTSGIQLLVSDRAHLRTAHGAALAGELPRPVLGCK 376

b 379 VAVDRLPSELAAPNAPRGQPLPHQCS-IHLNCEDTLLHQAEPDAMDGLPSHRPVIELC 437
Y 377 SPSGIDPTIAPAGRHQVTLWSQWQYRLSGHRDW-ASVAEABADRIUGEMEAFAFGFTDS 435
b 438 IPSSLDPTIAPPGCHVWSLFTQYTYTLGGKAWDEQERDAYADRVPCIEVYAPGFDS 497
Y 436 VLDRFIQTPRDIESBLGIMGNVHVENSLDQMLWRPLPELSGHRVPGADGLYLITGAST 495
b 498 VVGRDILTPDLERIFGLPGGNIFFCAMSLDQLYFTRFVPLHSGVRCF-LQGLYLCSGA 556
Y 496 HPGGVSGSGSRAARIALSD 516
c 557 HPGGVGMAAGRNAHVAFRD 577

35ULT 6
357177
ABB77177 standard; protein; 581 AA.
ABB77177;
08-OCT-2002 (first entry)
Human oxidoreductase protein OXIRED-8.
Human; oxidoreductase; OXIRED; cytostatic; neuroprotective; cardiant;
immunosuppressive; antiinflammatory; neuroleptic; endocrine; cancer;
neurological disorder; cardiovascular disorder; autoimmune disorder;
inflammatory disorder; metabolic disorder.

Homo sapiens.

WO200226951-A2.

04-APR-2002.

28-SEP-2001; 2001WO-US030656.

29-SEP-2000; 2000US-0237101P.

06-OCT-2000; 2000US-0236482P.

27-OCT-2000; 2000US-0244024P.

01-DEC-2000; 2000US-0250805P.

(INCY-) INCYTE GENOMICS INC.

Baughn MR, Yue H, Tribouley CM, Lee BA, Hafalia AJA;

Sanjanwala MS, Yao MG, Ding L, Elliott VS;

WPI; 2002-362493/39.

N-PSDB; ABL55865.

Human oxidoreductase polypeptide, useful in the diagnosis and treatment

of disorders associated with its activity, e.g., cancer and autoimmune

disorders.

Claim 1; SEQ ID NO 8; 117pp; English.

The sequence represents a novel human oxidoreductase polypeptide of the invention, designated OXIRED-8. The proteins of the invention have cytostatic, neuroprotective, cardiant, immunosuppressive, antiinflammatory, neuroleptic, and endocrine activity. The polypeptide is used in the diagnosis and treatment of disorders associated with decreased or increased expression or activity of oxidoreductases. Such disorders include cancer, cell proliferative disorders, neurological, cardiovascular, autoimmune/inflammatory and metabolic disorders. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Sequence 581 AA;

Query Match 22.6%; Score 624.5; DB 5; Length 581;
Best Local Similarity 33.0%; Pred. No. 2.e-43;

Matches 185; Conservative 81; Mismatches 226; Indels 69; Gaps 18;
QY 6 DAVVGSNNALVSAAYLAREGNSVEVLEKUTVLGAVSTVERFFGYKVDROSSAHLMIK 65
DB 36 DAVVIGAGHNGVAAYLQRLGVNTAVFERRHVIIGGAATEIIIPGFKPSRASYLLSLR 95
QY 66 HSGIIEELGLGAHGLRYIDCDPMATFAPPAPGTDGPG-----IVFHRDLDAFCQSIERAC 119
DB 96 PQ-IYTDLELKHGLRLHLRNYSFTPLM--EFGAGSKVPRCLLGTDAENQKQIAQ-F 151
QY 120 GTKODAYRRFVAVMSERSRHVMKAFSTPTG-----SNLIGAFGLATARG----- 166
DB 152 SKQDAQVFPKYEFEFMRHLAIDLPLDAAAPQHGSLQMRSLSTLKPLLKAGRI 211
QY 167 -NSELSPQF---LAPCDALIDYRDESEALKALAW---FGAOSGPPMSEPGTAPMVGFAA 219
DB 212 LGAQLPRYIEVLTAPITKVLQWFESEELKATLATDAVIGAMTSP--HTPGS---GYVL 265
QY 220 LMEVLPP-----GRAVGGSGALSAALASRMVADGATVALGDGVTSIRNSNEWT--V 269
DB 266 LHHVMGLGEGMQGAMGYVQSGMGALSDAIASSATTHGASIFTEKTVAKVQVNSECVQGV 325
QY 270 TTESGREVHARKVIACCH-----ILTTLDLNGGFDRTTLDHWRKIRVOPGIGAVLR 323
DB 326 VLEDGTEVRSKMVLNNTSPQITFLKTPQEWLPEFLEIRISQLOTRSPV-----TKIN 378
QY 324 LATSALPSY-----RGDATTRETSGLQLVSDRAHLTAHGAALAGELPPRPVAVLGM 376
DB 379 VAVDRLPSELAAPNAPRGQPLPHQCS-IHLNCEDTLLHQAEPDAMDGLPSHRPVIELC 437
QY 377 SPSGIDPTIAPAGRHQVTLWSQWQYRLSGHRDW-ASVAEABADRIUGEMEAFAFGFTDS 435
DB 438 IPSSLDPTIAPPGCHVWSLFTQYTYTLGGKAWDEQERDAYADRVPCIEVYAPGFDS 497
QY 436 VLDRFIQTPRDIESBLGIMGNVHVENSLDQMLWRPLPELSGHRVPGADGLYLITGAST 495
DB 498 VVGRDILTPDLERIFGLPGGNIFFCAMSLDQLYFTRFVPLHSGVRCF-LQGLYLCSGA 556
QY 496 HPGGVSGSGSRAARIALSD 516
DB 557 HPGGVGMAAGRNAHVAFRD 577

RESULT 7

ADC31532

ID ADC31532 standard; protein; 504 AA.

XX AC ADC31532;

XX AC ADC31532;

XX DT 18-DEC-2003 (first entry)

XX DE Human novel polypeptide sequence, SEQ ID NO:1614.

XX KW Human; diagnostic; drug screening; forensics; gene mapping;
XX KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
XX KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
XX KW ulcers; osteoporosis; autoimmune disease; cancer;
XX KW molecular weight marker; food supplement; antiparkinsonian; noctropic;
XX KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
XX KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
XX KW gene therapy; chromosome 10.

XX OS Homo sapiens.

XX PN WO2003029271-A2.

XX PD 10-APR-2003.

XX PF 24-SEP-2002; 2002WO-US030474.

XX PR 24-SEP-2001; 2001US-0324631P.

XX PA (HYSE-) HYSEQ INC.

contig sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a human contig-encoded polypeptide sequence used in an example of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 322 AA;

Query Match 11.0%; Score 305.5; DB 7; Length 322;
Best Local Similarity 30.5%; Pred. No. 4.9e-17;
Matches 101; Conservative 34; Mismatches 117; Indels 79; Gaps 11;
227 GRVGGGSAALASRAVMVDGATVALGDGVTISIRNSNHW- -VTTESGREYHARKVIA 284
26 GYVQGGGALSADASSATTHGASIFTEKTAKVQVNSEGCVGVLEDGTEVRSKWVLS 85
285 GCHILTLDLGNGGFFRTTLDHRRKIRVPGGICAVLRATSAALP-SYRGDATTRESIS 343
86 NTSPQIT- - - - -FLKLTPOEW- - - - -LPEFLERISOLDTRS 117
344 GQLLVSDRAHLRTARGAALA- - - - -GELPPPAVLGMSFGIDETIAP- - - - -AGR 390
118 PVTKINXEAR- - - - -HIAALSPLCHLSEKPGWQHLSHLLHCPDLQVSPCLVRSGR 173
391 HQVTLMSQWQYRLSG- - - - -HDMASVAEAEADRIVGEM 425
174 ROAAQPS-WRPPMLFGASRCPIITNAPSTXVTKTSPSIFRLKMPWMACTVFDCL- - - - -228
426 ZAFAPGFTDVLDFIOTFRDIESELGMIGNVHVMESLDQMLWRPLPELSCHRVGA 485
229 EYVAGFKDSVVGADILTPDLERIFGLPGGNIIFHCAMSLDQLYFAPVPLHSGIRCP-L 287
486 DGLYLTGASTHPGGVSGASGSAARIALSD 516
288 QGLYLCGSAHPGGVGMGAGENAARVAFRD 318

SUIT 9
AG90287
> AAG90287 standard; protein; 471 AA.

AAG90287;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 4041.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-00127668.

16-DEC-1999; 99JP-00377484.

07-APR-2000; 2000JP-00159162.

PR 03-AUG-2000; 2000JP-00280988.
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR N-PSDB; AAH65506.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analyzing expression profile or pattern of a gene and identifying homologous gene.
Claim 17; SEQ ID NO 4041; 246pp - Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analyzing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from Coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

XX Sequence 471 AA;

Query Match 10.7%; Score 296.5; DB 4; Length 471;
Best Local Similarity 25.9%; Pred. No. 4.6e-16;
Matches 141; Conservative 61; Mismatches 224; Indels 119; Gaps 23;

QY 7 AVVVGSGHNAVSAYLAREGNSVEVLEKOTVLGAVSTVERF-PGYKVDGSSAHLMI 65
DB 13 AVVVGSGNGUHTTAAVLAKAGQVDVYEAAPTGAARSESVLGEETISDLGAHGFEGV 72
QY 66 HSGIIEELGAGHGLRYIDCDPMAFAPPA- - -PGTDGFGIVFHRDLDTQCSTIERACGTX 122
DB 73 ASPAPHYLGLDGHLE- - - - -WAYSPPFAWHPDYGRAGLETSLPETAKKL- - - - -119
QY 123 DADAVRRFVAVNSERSRHVKAFSTPTPGNSLIG- - - - -AFGLLATARGNSE 169
DB 120 GPDA-RRWKNLHQGLTKNIDKHL- - - - -ANLGPVLKWAHPAHPMAKPGPFA- - - - -165
QY 170 LSRQFLAPGDALDDEYFDSEALKAALAWFGAQSGPPMSEPCAPMVGFPAALMHVLPFGR- 228
DB 166 - - - - -LLPAKLASAAFEETEARSIFICSAMHSVTPPHKPMWASL- - - - -GLLFGALGWSRG 217
QY 229 - - - - -AVGGGSAALASRAVMVDGATVALGDGVTISIRNSNHWTTTESGREYHARKVIAG 285
DB 218 WFAVVGSGRIVDALWNVHHGGTIHCDSDISL- - - - -SQFRDQTD- - - - -ILIN 264
QY 286 CHILTLDLIG- - - - -NGGPDRTTLDHRRKIRVPGGICAVLRATSAALP- - - - -SYRGDATT 339
DB 265 QTPSQVLKLTGDLNAGLPQ-RMSTNKH- - - - -GPSSYKYDYLDDGPIWSPNQVQAATT 319
QY 340 ESTGSLQLLVSDRAHLRTAHGAALAGBELPPPAVLGMSFGIDETIAPAGRHQVTLMSQW 399
DB 320 H- - - - -VGGSSSEIAPAEAEVAARMPERPFITLCCQQVADPSRAREGSHVY- - - - -WA- - - 368
QY 400 QPYRLSGHRDWAASAEABDR- - - - -IVGMEAFAPGTDSDVLDRFICTPREDBS-BLGM 453
DB 369 - - - - -YAVPRGPFVDRRAALLITAQIERFAPGFRDRIHVHSDVNTNAEDLEAWPNL 418
QY 454 IGGNVHVMESLDQMLWRPLPELSCHRVFGADGLYLTGASTHPGGVSGASGSAARIA 513
DB 419 VGGDI- - - - -TAGSALLRRWPTKIGECT- - - - -YMASASNADPGGVGMGPGWAAQAV 466
QY 514 LSDSR 518
| : |

0000334-AZ.

489 YLTGASTHPOGGV-----SG--ASGRSAARIALS-----DSRRGKASQMMR 527
561 YCVGDSFCPOGGVIAVAFSGWMCARVAADLGFPEKSDVLDLALLRLGLWR 612

RESULT 13
AB85728
D AAB85728 standard; protein; 548 AA.

C AAB85728;
T 29-OCT-2001 (first entry)
E Enzyme involved in carotenoid biosynthetic pathway.
X Carotenoid; enzyme; catalyst; geranylgeranylpyrophosphate, phytoene;
W farnesylpyrophosphate.
S Brevibacterium flavum.
X JP2001149077-A.

D 05-JUN-2001.

F 24-NOV-1999; 99JP-00333395.

X 24-NOV-1999; 99JP-00333395.

X (MITU) MITSUBISHI CHEM CORP.

A WPI; 2001-499379/55.

R N-PSDB; AAB76025.

X New DNA encoding enzyme of new carotenoid biosynthetic path, useful for

I preparation of carotenoid.

S Claim 5; Page 23-25; 33pp; Japanese.

X The invention relates to enzymes involved in a carotenoid biosynthetic
C pathway and DNA sequences encoding the polypeptides. One polypeptide has
C the activity of catalyzing a reaction forming geranylgeranylpyrophosphate
C from farnesylpyrophosphate. Another polypeptide has an activity of
C catalyzing a reaction forming phytoene from geranylgeranylpyrophosphate.
C The DNA can be used for the preparation of carotenoid. The present
C sequence represents an enzyme of carotenoid biosynthetic pathway

X Sequence 548 AA;

Query Match 8.0%; Score 221.5; DB 4; Length 548;
Best Local Similarity 25.8%; Pred. No. 1.1e-09;
Matches 141; Conservative 66; Mismatches 236; Indels 103; Gaps 27;

Y 7 AVVVGSHNALSAAYLAREGWSVELEKTVLGGAVTVE--RPPGYKVDKSGSAHLMI 64
b 14 AVVVGAGVAGLATSALLARDGQVTVLEKTVDVGGRAGSLSEIDPGRWDTGSWILMP 73
Y 65 RSGIIEELGLAHGURYIDCFWAFAPAPPTDGGIVF---HRDL-----ATC 112
b 74 EAFDHPFAL-FCARTSDYLD-----LVELTPGYR----VFGTHDAVDVPGREKAIALF 123
Y 113 OSTERACGTK-----DADAYRRFVAVMSRSRHVMKAFST--PPTGSLNIGAFGGIAT 163
b 124 ESTEPAGAKLGNLYDSADAYD--IAI-----DRFLYNNFSTLGPLHRDVLTRAGRLP- 176
Y 164 ARGNSELSQFLAPGDALDEYFDSEALKAAALANFAGSGPPM---SEPGTAPMVGPAAL 220
b 177 ----SLTTRSL-----QKYVNSQFSSPVLKQLTY-----PAVFLSSRPTTTPSM-VHLM 221
Y 221 MHV-LPPG--RAVGGGALSALASRMVADGATVAGDGVTSIRENSNHWITVTTESGREV 277
b 222 SHYDLVQGVKYPGGGTAVVNAHLQALENGVEFOLDSEVISINTASSRGNTSATGVSFL 281
Y 278 HARKVIAGCHILTTLD---LLNGGFFDRTTLDHWRKIRV-----GPGIGAVLRL- 324

Db 282 HNRKV-----QNLDAVLWSAGDLRHTENNLLPRELATYPERYWSNPNPGIGAVLILL 334
Qy 325 -ATSAIPSY-RGDATTRBSTSGIQLLVSDRAHLRTAAGALAGELPPRPAVLGMSFGID 382
Db 335 GVKGELPQLDHHNLFFESDWTDDFAVVFGPQPTRPHNAS-----NSIYVSKFSTSE 386
Qy 383 PTIAPAGRHOVTLWSQWQPYRLSGHRD-WASVAEAEADRIVGE-----MEAPAPGFTDS 435
Db 387 DGVAPAGYENLFLIPAKASSSIGHGDAYWQSASASVETIASHAINQIATQAGIDPLDTR 446
Qy 436 VLDRFTQTPRDIASELGMIGNVHVMESLDQWMLWRPLPELSGHRVPGADGILYLTGAST 495
Db 447 IVVKRTIGPADFHRYHSWVGSGALGPAHTLROSAFLR--GRNSSRKV----DNLFYSGATT 501
Qy 496 HPGGGV 501
Db 502 VPGVGI 507

RESULT 14

AAB76640

ID AAB76640 standard; protein; 548 AA.

XX AAB76640;

XX 11-APR-2001 (first entry)

XX Corynebacterium glutamicum MCT protein SEQ ID NO:262.

XX Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT;
XX membrane construction and membrane transport protein; petroleum spill;
XX hydrocarbon degradation; gram positive aerobic bacterium; marker;
XX identification; microorganism; fine chemical production; transformation;
XX genome mapping; genetic engineering.

XX Corynebacterium glutamicum.

XX WO200100805-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000MO-IB000926.

XX 25-JUN-1999; 99US-0141031P.

XX 08-JUL-1999; 99DE-01031454.

XX 08-JUL-1999; 99DE-01031478.

XX 08-JUL-1999; 99DE-01031563.

XX 08-JUL-1999; 99DE-01032122.

XX 08-JUL-1999; 99DE-01032124.

XX 08-JUL-1999; 99DE-01032125.

XX 08-JUL-1999; 99DE-01032128.

XX 08-JUL-1999; 99DE-01032180.

XX 08-JUL-1999; 99DE-01032182.

XX 08-JUL-1999; 99DE-01032190.

XX 08-JUL-1999; 99DE-01032191.

XX 08-JUL-1999; 99DE-01032209.

XX 08-JUL-1999; 99DE-01032212.

XX 08-JUL-1999; 99DE-01032227.

XX 08-JUL-1999; 99DE-01032228.

XX 08-JUL-1999; 99DE-01032229.

XX 08-JUL-1999; 99DE-01032230.

XX 14-JUL-1999; 99DE-01032297.

XX 14-JUL-1999; 99DE-01033005.

XX 14-JUL-1999; 99DE-01033006.

XX 27-AUG-1999; 99DE-01040764.

XX 27-AUG-1999; 99DE-01040765.

XX 27-AUG-1999; 99DE-01040766.

XX 27-AUG-1999; 99DE-01040830.

XX 27-AUG-1999; 99DE-01040831.

XX 27-AUG-1999; 99DE-01040832.

XX 31-AUG-1999; 99DE-01040833.

XX 31-AUG-1999; 99DE-01041378.

x Claim 20; Page 538-540; 1119pp; English.
 S AAF57743 to AAF68080 encode the Corynebacterium glutamicum membrane
 C construction and membrane transport (MCT) proteins given in AAB76510 to
 C AAB76847. The MCT nucleic acids and proteins are useful in the
 C identification of microorganisms which can be used to produce fine
 C chemicals, for modulating fine chemical production in C. glutamicum or
 C related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 C identification of C. glutamicum or related bacteria, as reference points
 C for mapping C. glutamicum genome, and as markers for transformation.
 C AAF68082 and AAF68082 represent sequencing primers which are used in an
 C example from the present invention
 x
 2 Sequence 548 AA;
 Query Match 7.9%; Score 219.5; DB 4; Length 548;
 Best Local Similarity 25.8%; Pred. No. 1.6e-09;
 Matches 139; Conservative 68; Mismatches 243; Indels 89; Gaps 26;
 7 AVVVGSHNALVSAAVLAREGWSVEVLEKDTVLGGAVSTVE--RPGYKYDRGSSAHLM 64
 14 AVVIGAGVAGLATSALLARDGQVTLKNTDVGGRAGSLEISGPGFRMDTGPSWILMP 73
 65 RNSGIIIEELGAGHLYID--CDPW--APAPPAGTDPGIVFHRDLDTATCOTIERAC 119
 74 EAPDHPFAL--FGACTSDVLDVLELTPGVRFSGTHDAVDP--TCREAIAPESIEPGA 130
 120 GTK-----DADAYRRFVAVWSESRHVMKAPST--PPTGSLNIGAPGGLATARGNSEL 170
 131 GAKLGNLYDSAADAYD--IAI----DRFLYNNFSTGLPLLRDVLTRAGRLF-----SL 179
 171 SRQFLAPGDALLDEYFDSBALKAALAWFGAQSGPPM--SEPCTAPMVGFALMHV-LPP 226
 180 TRSL-----QKYVNSQFSSPVLRQILTY-----PAVFLSSRPPTTTPSM-YHLMSTDLVQ 228
 227 G--RAVGGSGALSAAALASRMAYDGTATVAGDGVTSIRNSNHWITVTSSGREVHARKYIA 284
 229 GVKYPIGGFTAVVNALHQLALENGVEFQIDSEVISINTASSRGNTSANGVSLLEHKKV-- 286
 285 GCHILLTLB---LLGNGGDFRTTLDHNRKRV-----GPGIGAVLEL--ATSALP 330
 287 -----QNLDAILVVSAGDLHHTENLLPRELTPERYWNSNRNPGIGAVLILLGVKGELP 341
 331 SY-RGDATTRESTSGILQLIVSDRAHLRTAHGAALAGELPPRPVILGMSFGIDPTIAPAG 389
 342 QLDHNLFTSEDTDDFAVVDGPGQLTRPHNAS-----NSIYVSKPSTSESGVAPAG 393
 390 RHQVTLMSQWQPYRLSGHRD--WASVAEAEADRIUGE-----MEAFPGFTDSDVLDRIQ 442
 394 YENLFLVLIPTKASSSTIGHGDATYMQASASVETIASHAINQIATQAGIPDLTDRIVVKRTI 453
 443 TPRDIESELGMIGGMVHMVEMSLDQMLWRPLPELSGHRVPCADGLYLTGASTHPGGGV 501
 454 GPADFPHRYHSWVGSGALGPAHTLRQSAFLR--GRNSSRKV---NNLFYSGATTVGVGI 507

Search completed: February 29, 2004, 14:44:12
 Ob time : 66.9999 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 14.3897 Seconds
(without alignments)
3837.172 Million cell updates/sec

title: US-09-941-947a-36

affect score: 956
sequence: 1 MLWLNALIVFTVVGMEWV.....ARSGAARDQGVDTSSSGK 175

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_ARCHES.*
- 2: SP_BACTERIA.*
- 3: SP_FUNGI.*
- 4: SP_HUMAN.*
- 5: SP_INVERTEBRATE.*
- 6: SP_MAMMAL.*
- 7: SP_MHC.*
- 8: SP_ORGANELLE.*
- 9: SP_PHAGE.*
- 10: SP_PLANT.*
- 11: SP_RODENT.*
- 12: SP_VIRUS.*
- 13: SP_VIRUS.*
- 14: SP_UNCLASSIFIED.*
- 15: SP_RVIRUS.*
- 16: SP_BACTERIAP.*
- 17: SP_ARCHAEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	956	100.0	175	2 Q8GCR9	Q8GCR9 Pantoea ste
2	822	86.0	175	2 Q8VUJ3	Q8VUJ3 Pantoea agg
3	492	51.5	162	2 Q8RLH6	Q8RLH6 Paracoccus
4	487	50.9	169	2 P94792	P94792 Flavobacter
5	468	49.0	155	16 Q88HV7	Q88HV7 pseudomonas
6	271	28.3	151	17 Q97UT7	Q97UT7 sulfobolus
7	271	28.3	303	10 Q9LTC0	Q9LTC0 arabidopsis
8	270	28.2	308	10 Q9FV13	Q9FV13 narcissus p
9	270	28.2	322	10 Q9SPK6	Q9SPK6 haematococ
10	263.5	27.6	311	10 Q9FTN2	Q9FTN2 citrus unsh
11	263.5	27.6	311	10 Q9FUB9	Q9FUB9 citrus unsh
12	263	27.5	299	10 Q8LKV3	Q8LKV3 vitis vinif
13	262.5	27.5	305	10 Q8VXP2	Q8VXP2 crocus sat
14	261	27.3	315	10 Q49815	Q49815 capsicum an
15	260.5	27.2	316	10 Q49814	Q49814 capsicum an
16	257.5	26.9	314	10 Q9S6Y0	Q9S6Y0 lycopersico

17	257	26.9	296	10	Q84U11	Q84U11 haematococ
18	256.5	26.8	148	2	Q7WT73	Q7WT73 marine bact
19	254	26.6	309	10	Q8S6V1	Q8S6V1 lycopersico
20	253	26.5	294	10	Q36237	Q36237 arabidopsis
21	253	26.5	309	10	Q9FV41	Q9FV41 tegetes ere
22	253	26.5	310	10	Q9S228	Q9S228 arabidopsis
23	247	25.8	292	10	Q8LN34	Q8LN34 oryza sativ
24	245	25.6	309	10	Q7XQ82	Q7XQ82 oryza sativ
25	231	24.2	127	10	Q8RY53	Q8RY53 sandersonia
26	98.5	10.3	410	16	P74167	P74167 synechocyst
27	96.5	10.1	324	16	Q8XZ21	Q8XZ21 raietonia s
28	96	10.0	315	16	Q98MP4	Q98MP4 rhizobium l
29	93.5	9.8	355	16	Q986V0	Q986V0 rhizobium l
30	93	9.7	418	10	Q8W3L3	Q8W3L3 chlorella v
31	92	9.6	408	2	Q8GCR9	Q8GCR9 streptomyc
32	91.5	9.6	391	10	Q9ZPP7	Q9ZPP7 perilla fru
33	91	9.5	313	16	Q8UIH7	Q8UIH7 agrobacteri
34	90.5	9.5	444	10	Q8H1X8	Q8H1X8 betula verr
35	89	9.3	163	10	Q9SX07	Q9SX07 arabidopsis
36	88	9.2	301	16	Q9ISW4	Q9ISW4 pseudomonas
37	87.5	9.2	302	10	Q93Y93	Q93Y93 brassica ra
38	87	9.1	335	16	Q8ZG25	Q8ZG25 streptomyc
39	87	9.1	600	5	Q961U9	Q961U9 drosophila
40	87	9.1	674	5	Q9VIK2	Q9VIK2 drosophila
41	86.5	9.0	104	10	Q9XHS5	Q9XHS5 brassica na
42	86.5	9.0	108	10	Q9XHS6	Q9XHS6 brassica ol
43	86.5	9.0	429	10	Q9M4D4	Q9M4D4 brassica ju
44	86.5	9.0	435	10	Q8W4M9	Q8W4M9 arabidopsis
45	86	9.0	445	16	Q8PQ16	Q8PQ16 xanthomonas

ALIGNMENTS

RESULT 1

Q8GCR9 PRELIMINARY; PRT; 175 AA.
 AC Q8GCR9
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Beta-carotene hydroxylase.
 GN CRTZ.
 OS Pantoea stewartii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=66269;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 8200;
 RA deSouza M.L., Kollmann S.R., Schroeder W.A.;
 RT "Carotenoid Biosynthesis (WO 02/079395 A2)";
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY166713; AAN85601.1; -
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0008152; F:metabolism; IEA.
 DR InterPro; IPR005596; Carotene hydrox.
 DR InterPro; IPR006087; Sterol desat.
 DR Pfam; PF03897; Carotene hydrox. 1.
 SQ SEQUENCE 175 AA; 19884 MW; D7F1069562D83054 CRC64;

Query Match 100.0%; Score 956; DB 2; Length 175;
 Best Local Similarity 100.0%; Pred. No. 4.8e-89;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLWLNALIVFTVVGMEVVAALAHKYLIMHGWGWSHSHPRKGAPEVNDLVAVVFAI	60
Db	1	MLWLNALIVFTVVGMEVVAALAHKYLIMHGWGWSHSHPRKGAPEVNDLVAVVFAI	60
QY	61	VSIALIYFGSTGIMPLQWIGAGMTAYGLLYFMVHDGLVHQWPPRYIPRKGYLRLVNAH	120
Db	61	VSIALIYFGSTGIMPLQWIGAGMTAYGLLYFMVHDGLVHQWPPRYIPRKGYLRLVNAH	120

121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175
121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175

RESULT 2

Q8VUJ3 PRELIMINARY; PRT; 175 AA.
Q8VUJ3, 2002 (TrEMBLrel. 20, Created)
01-MAR-2002 (TrEMBLrel. 23, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
CrtZ protein.
N CRTZ
S Pantoea agglomerans pv. milletiae.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=182454;
N N
P SEQUENCE FROM N.A.
A Kamiunten H., Hirata R.;
T "Isolation and characterization of carotenoid biosynthesis genes from
T Pantoea agglomerans pv. milletiae Wist 801.";
L Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; AB076662; BAB79605.1; -
R GO: GO:0003824; F:catalytic activity; IEA.
R GO: GO:0008152; P:metabolism; IEA.
R InterPro; IPR005596; Carotene_hydrox.
R InterPro; IPR006087; Sterol_desat.
R Pfam; PF03897; Carotene_hydrox; 1.
R PROSITE; PS00018; EF_HAND; 1.
Q SEQUENCE 175 AA; 20222 MW; 4849DDAC61167845 CRC64;

Query Match 86.0%; Score 822; DB 2; Length 175;
Best Local Similarity 82.3%; Pred. No. 1.9e-75;
Matches 144; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

1 MLEINNALIVFVTVGMEVVAALAHKYIMHG-WGHWLSSHH-EPKGAPEVNDLYAVVFAI 60
1 MLEINNALIVFVTVGMEVVAALAHKYIMHG-WGHWLSSHH-EPKGAPEVNDLYAVVFAA 60
61 VSALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAH 120
61 LSILLIYVAGSTGWPWQWIGAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAH 120
121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175
121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175

RESULT 3

Q8RLH6 PRELIMINARY; PRT; 162 AA.
Q8RLH6, 2000 (TrEMBLrel. 13, Created)
01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
Carotene hydroxylase.
N CRTZ
S Paracoccus marcusii.
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
C Rhodobacteraceae; Paracoccus.
X NCBI_TaxID=59779;
N N
P SEQUENCE FROM N.A.
A Harker M., Hirschberg J.;
T "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
T Mh1.";
L Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
R EMBL; Y15112; CAB56060.1; -
R GO: GO:0003824; F:catalytic activity; IEA.
R GO: GO:0008152; P:metabolism; IEA.
R InterPro; IPR005596; Carotene_hydrox.

DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
SQ SEQUENCE 162 AA; 18882 MW; 798891BFF8743999 CRC64;

Query Match 51.5%; Score 492; DB 2; Length 162;
Best Local Similarity 57.4%; Pred. No. 5.3e-42;
Matches 93; Conservative 20; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFVTVGMEVVAALAHKYIMHG-WGHWLSSHH-EPKGAPEVNDLYAVVFAIVSIA 64
3 NFLIVVATLVVMEITAYSVHRWIMHGFLGWGHWKSHHEDHLEKNDLYGLVFAIVATV 62
65 LIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAHRM 122
63 LFTVVG-WIAPVLMWIALGMTVYGLIYFVLDHGVHQRWPFYIPRKGYARBLQAHL 120
123 HHAVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDE 164
121 HHAVEGRDHCVSFGFTYAPPVVDKLDKX-----TSGVLRAE 157

RESULT 4

P94792 PRELIMINARY; PRT; 169 AA.
AC P94792;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DN B-carotene hydroxylase.
CN CRTZ
OS Flavobacterium sp. ATCC 21588.
OC Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;
OC Flavobacteriaceae; Flavobacterium.
OX NCBI_TaxID=50286;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL1534;
RX MEDLINE=97186694; PubMed=9034310;
RA Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,
RA van Loon A.P.;
RT "Isolation and characterization of the carotenoid biosynthesis genes
RT of Flavobacterium sp. strain RL1534.";
RL Gene 185:35-41(1997).
DR EMBL; U62808; AAC44852.1; -
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
DR PROSITE; PS00018; EF_HAND; 1.
SQ SEQUENCE 169 AA; 19282 MW; 85334038DB3DAAB1 CRC64;

Query Match 50.9%; Score 487; DB 2; Length 169;
Best Local Similarity 51.5%; Pred. No. 1.8e-41;
Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;

5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGHWLSSHH-EPKGAPEVNDLYAVVFAIV 61
4 WAAILLVTLVTAAMELTAYSVHRWIMHGFLGWGHWKSHHEDHLEKNDLYGLVFAIV 63
62 SIALLIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAHR 121
64 SIVLPAIGAMGSDLAWLAVGVTCYGLIYFLDGLSVHGRWPFYIPRKGYLRLYMAHR 123
122 MHAVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSS 172
124 MHAVHAGRENCVSPGFINAPSVDSLKAEK-----RSGALLKOREGADNRT 169

RESULT 5

Q88HV7

```

RESULT 7
ID Q9LTG0 PRELIMINARY; PRT; 303 AA.
AC Q9LTG0;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative beta-carotene hydroxylase.
DS AF5652570.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosoids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
[1]_TaxID=3702;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

```

structural analysis of Alaudinopsis caroliniana chromosome 3. XI.
Submitted (APR-1999) to the EMBL/GenBank/DDJB databases.

[2]
SEQUENCE FROM N.A.
RN Yamada K., Banb J.J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Mizutani M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Eckert J.R., Theologis A.,
RT "Full length cDNA of gene At5g52570 (GI:15237204)".
RL Submitted (JAN-2002) to the EMBL/GenBank/DDJB databases.

[3]
SEQUENCE FROM N.A.
RN Yamada K., Banb J.J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,


```
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.,
RA "Arabidopsis Open Reading Frame (ORF) Clones.",
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025606; BAA98075.1; -
DR EMBL; AY074394; AAL67090.1; -
DR EMBL; AY117225; AAM51300.1; -
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
DR SEQUENCE 303 AA; 33777 MW; C7B8C9BF8B8E34CB CRC64;

Query Match 28.3%; Score 271; DB 10; Length 303;
Best Local Similarity 42.0%; Pred. No. 2.9e-19;
Matches 55; Conservative 26; Mismatches 44; Indels 6; Gaps 3;

2y 15 VGMEVVAALAHKYIMHGWGWSHSHHEPRKGAPEVNDLYAVVFAIVSIALIYFG--STG 72
2b 145 VGMEFWARWAHRAHWDLSLNWCHESHKHPREGAPELNDVFAITNAVPAIGLIIYGFINKG 204
2y 73 IWPQLQWAGCM--TAYGLLFVMDHGLVHQRWPRYIPRKGYLKRLVMAHRMHAHVRGKE 130
2b 205 LVPLGCLGAGLGTIFMGCMAYMFVHDGLVHKRFPVGPPIANVPYILRKVAARHQLHTDKPKG 264
2y 131 GCVSFGFLYAP 141
2b 265 --VPYGLFLGP 273

RESULT 8
ID Q9FY13 PRELIMINARY; PRT; 308 AA.
AC Q9FY13;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-carotene hydroxylase.
DC Narcissus pseudonarcissus (Daffodil).
DC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
DC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
CC Narcissus.
DX NCBI_TaxID=39639;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Flower;
RA Schaub P., Beyer P., Al-Babili S.;
RT "A cDNA encoding beta-carotene hydroxylase from Narcissus
pseudonarcissus."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278882; CAC06712.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
DR SEQUENCE 308 AA; 34849 MW; 2782E3027252D30C CRC64;

Query Match 28.2%; Score 270; DB 10; Length 308;
Best Local Similarity 43.5%; Pred. No. 3.7e-19;
Matches 57; Conservative 22; Mismatches 46; Indels 6; Gaps 3;

2y 15 VGMEVVAALAHKYIMHGWGWSHSHHEPRKGAPEVNDLYAVVFAIVSIALIYFG--STG 72
2b 150 VGMEFWARWAHRAHWDLSLNWCHESHKHPREGAPELNDVFAITNAVPAISLLIYGFNRG 209
2y 73 IWPQLQWAGCM--TAYGLLFVMDHGLVHQRWPRYIPRKGYLKRLVMAHRMHAHVRGKE 130
2b 210 LVPLGCLGAGLGTIYGMAYMFVHDGLVHRRFPVGPPIADVPPYPRVAAHRIHT--EKF 267
2y 131 GCVSFGFLYAP 141
```

```
Db 268 NGVPYGLFLGP 278

RESULT 9
ID Q9SPK6 PRELIMINARY; PRT; 322 AA.
AC Q9SPK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Carotenoid hydroxylase (Fragment).
OS Haematococcus pluvialis.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Haematococcaceae; Haematococcus.
OX NCBI_TaxID=44745;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99453720; PubMed=10524195;
RA Linden H.;
RT "Carotenoid hydroxylase from Haematococcus pluvialis: cDNA sequence,
regulation and functional complementation.";
RL Biochim. Biophys. Acta 1446:203-212 (1999).
DR EMBL; AF162276; AAD54243.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR004048; Reg_chromat.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
DR PROSITE; PS00626; RCC1_2; 1.
FT NON TER 1
FT SEQUENCE 322 AA; 34986 MW; 45100BASECBA13EC CRC64;

Query Match 28.2%; Score 270; DB 10; Length 322;
Best Local Similarity 42.0%; Pred. No. 3.9e-19;
Matches 60; Conservative 20; Mismatches 49; Indels 14; Gaps 4;

Qy 8 LIVFTVTVGVMEVVAALAHKYIMHGWGWSHSHHEPRKGAPEVNDLYAVVFAIVSIAL 65
Db 160 LLLVGGALGEMVARYAHKAIHESPLGWLHSHHTPTGPFENDLFAINGLPAMLL 219
Qy 66 IYFGSTGIWPLQWIGA-----GMTAYGLLFVMDHGLVHQRWPRYIPRKGYLKRLYM 118
Db 220 CTF---GFLLPNVLGAACFGAGLGTIYGMAYMFVHDGLVHRRFPVGPPIAGLFYMKRLTV 276
Qy 119 AHRMHAHVRGKGCVSFGFLYAP 141
Db 277 ARLHRS--GRYGGAPWGMFLGP 297

RESULT 10
ID Q9FUN2 PRELIMINARY; PRT; 311 AA.
AC Q9FUN2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-carotene hydroxylase.
GN CHX1.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa; TISSUE=Fruit;
RA Kim I., Ko K., Kim C., Chung W.;
RT "Isolation of a cDNA encoding beta-carotene hydroxylase from Citrus.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF296158; AAG10793.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
```


Best Local Similarity 38.6%; Pred. No. 2.1e-18;
Matches 61; Conservative 29; Mismatches 55; Indels 13; Gaps 5;
Y 15 VGEVVAALAHKYMKGWGWHLSSHEPRKGAPEVNDLYAVVFAIVSIALIYFG--STG 72
b 146 VGEFEWARWAHRAHMAHSHWHESSHRPRGPFELNDVFAIINAVPALALNFGFFRG 205
Y 73 IWPLOWIGAGM--TAYGLLYFMVHDGLVHQRWPFYIPRKGYLKLYMAHRMHAHVGRKE 130
b 206 LFPCLCFGAGLITLFGIAYMFVHDGLVHRRFPVGPDIADVPYQFQVAAAHQIHSEK-FE 264
Y 131 GCVSFGFLYAPP-----LSKQATLRERHARSAA 161
b 265 G-VPYGLFPGKPELEIGGLKELEKEVSRIRKAYNNSA 301

RESULT 14

49815 PRELIMINARY; PRT; 315 AA.
D O49815
C O49815;
T 01-JUN-1998 (TREMELrel. 06, Created)
T 01-JUN-1998 (TREMELrel. 06, Last sequence update)
T 01-JUN-2003 (TREMELrel. 24, Last annotation update)
E Beta-carotene hydrolase.
S Capsicum annuum (Bell pepper).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C lamids; Solanales; Solanaceae; Capsicum.
X NCBI_TaxID=4072;
[1]
N SEQUENCE FROM N.A.
IP MEDLINE=98223452; PubMed=9555077;
IX Bouvier F., Keller Y., D'Harlingue A., Camara B.;
JA "Xanthophyll biosynthesis : molecular and functional characterization
of carotenoid hydroxylases from pepper fruits (Capsicum annuum L.).";
IL Biochim. Biophys. Acta 1391:320-328(1998).
JR EMBL; Y09225; CAA70427.1; --
JR GO; GO:0016787; F:hydrolase activity; IEA.
JR GO; GO:0008152; P:metabolism; IEA.
JR InterPro; IPR005596; Carotene_hydrox.
JR InterPro; IPR006087; Sterol_desat.
JR Pfam; PF03897; Carotene_hydrox; 1.
CW Hydrolase.
SQ SEQUENCE 315 AA; 35426 MW; F68792134297E3B4 CRC64;

Query Match 27.3%; Score 261; DB 10; Length 315;
Best Local Similarity 35.4%; Pred. No. 3.1e-18;
Matches 54; Conservative 25; Mismatches 52; Indels 6; Gaps 3;
DY 9 IVFVVVGEVVAALAHKYMKGWGWHLSSHEPRKGAPEVNDLYAVVFAIVSIALIYF 68
DB 155 LAFGAAIGMEYWARWAHRAHMAHSHWHESSHRPRGPFELNDVFAIINAVPALAFPSF 214
DY 69 G--STGIWFLQWIGAGM--TAYGLLYFMVHDGLVHQRWPFYIPRKGYLKLYMAHRMHH 124
DB 215 GFNFKGLIPGICFGAGLITVFGMAFMVHDGLVHRRFPVGPDIADVPYQFQVAAAHQHLH 274
DY 125 AVRGKGCVSFGFLYAP 141
DB 275 S--DKFCGVPGFLGP 289

RESULT 15

249814
ID O49814 PRELIMINARY; PRT; 316 AA.
AC O49814;
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Beta-carotene hydroxylase 2 (fragment).
DS Capsicum annuum (Bell pepper).
DC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
DC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC lamids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98223452; PubMed=9555077;
RA Bouvier F., Keller Y., D'Harlingue A., Camara B.;
RT "Xanthophyll biosynthesis : molecular and functional characterization
of carotenoid hydroxylases from pepper fruits (Capsicum annuum L.).";
RL Biochim. Biophys. Acta 1391:320-328(1998).
DR EMBL; Y09225; CAA70888.1; --
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
FT NON_TER 1
SQ SEQUENCE 316 AA; 35792 MW; F59395B6DA552329 CRC64;

Query Match 27.2%; Score 260.5; DB 10; Length 316;
Best Local Similarity 38.0%; Pred. No. 3.5e-18;
Matches 57; Conservative 28; Mismatches 52; Indels 13; Gaps 4;
OY 15 VGEVVAALAHKYMKGWGWHLSSHEPRKGAPEVNDLYAVVFAIVSIALIYFG--STG 72
DB 162 VGEFEWARWAHRAHMAHSHWHESSHRPRGPFELNDVFAIINAVPALALDYGFPHKG 221
OY 73 IWPLOWIGAGM--TAYGLLYFMVHDGLVHQRWPFYIPRKGYLKLYMAHRMHAHVGRKE 130
DB 222 LFPCLCFGAGLITVFGMAFMVHDGLVHRRFPVGPDIADVPYQFQVAAAHSLHS--EKF 279
OY 131 GCVSFGFLYAPP-----LSKQATLRER 153
DB 280 NGVPYGLFPGKPELEIGGLKELEKEVNR 309

Search completed: February 29, 2004, 14:51:08
Job time : 18.3897 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:27:18 ; Search time 2.94944 Seconds
(without alignments)
3089.496 Million cell updates/sec

tie: US-09-941-947A-36

fect score: 956

quence: 1 MLMTWNNALIVFTVVGMEV.....ARSGAARDQDGVDTSSSGK 175

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

al number of hits satisfying chosen parameters: 141681

nimum DB seq length: 0

ximum DB seq length: 200000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	892	93.3	175	1 CRTZ_PANAN	P21688 pantoea ana
2	654	68.4	176	1 CRTZ_ERWHE	Q01332 erwinia her
3	499	52.2	162	1 CRTZ_ALCSP	Q42622 alcaligenes
4	495	51.8	162	1 CRTZ_AGRAU	P54973 agrobacteri
5	101.5	10.6	383	1 F03E_BRANA	P48624 brassica na
6	93.5	9.8	386	1 F03E_ARATH	P48623 arabidopsis
7	87	9.1	934	1 CAPP_MYCLE	P46710 mycobacteri
8	86.5	9.0	435	1 F03D_ARATH	P48622 arabidopsis
9	86	9.0	447	1 F03C_SESIN	P48620 sesamum ind
10	84.5	8.8	404	1 F03C_BRANA	P48618 brassica na
11	83	8.7	350	1 LSS1_HUMAN	P27544 homo sapien
12	82.5	8.6	446	1 F03C_ARATH	P46310 arabidopsis
13	82	8.6	562	1 CYOB_BUCAR	P37543 buchnera ap
14	81.5	8.5	460	1 F03C_RICCO	P48619 ricinus com
15	80	8.4	659	1 CYOB_BUCAR	Q8K994 buchnera ap
16	78	8.2	370	1 COMM_EROME	Q24139 drosophila
17	78	8.2	551	1 QCRH_MYCLE	P35878 mycobacteri
18	77.5	8.1	168	1 LSPA_VIBCU	Q85878 vibrio vuln
19	77.5	8.1	410	1 NUOH_MYCTU	P30220 epstein-bar
20	76.5	8.0	332	1 UL95_EBV	P48621 glycine max
21	76.5	8.0	453	1 F03C_SOYBN	P34700 pediococcus
22	76.5	8.0	651	1 PPSA_PEDPE	P50939 rhodobacter
23	76.5	8.0	712	1 NUO8_RHOCA	Q50678 mycobacteri
24	76	7.9	542	1 YMB7_MYCTU	Q87889 vibrio para
25	75.5	7.9	169	1 LSPA_VIBBI	P15078 escherichia
26	75.5	7.9	701	1 CSTA_ECOLI	P7545 mus musculus
27	75	7.8	350	1 LSS1_MOUSE	P23477 bacillus su
28	75	7.8	1166	1 ADBA_BACSD	P72637 synechocyst
29	74	7.7	1032	1 YAS60_SVNT3	P26751 human cytom
30	72.5	7.6	431	1 UL78_HCMVA	Q9zel5 rickettsia
31	72	7.5	236	1 Y147_RICPR	Q8zch3 yersinia pe
32	72	7.5	510	1 NANT_VERPE	Q57491 haemophilus
33	71.5	7.5	471	1 Y872_HAEIN	

34 71.5 7.5 527 1 MEMA_METCA
35 71.5 7.5 654 1 NUOL_STROO
36 71.5 7.5 946 1 Y2T6_YEAST
37 70.5 7.4 452 1 WZYE_SALTI
38 70.5 7.4 452 1 WZYE_SALTI
39 70.5 7.4 452 1 WZYE_SALTI
40 70.5 7.4 452 1 WZYE_SALTI
41 69.5 7.3 171 1 LSPA_VIBCH
42 69.5 7.3 321 1 ALX_ECOLI
43 69.5 7.3 421 1 TRAB_AGRIS
44 69.5 7.3 424 1 FD6C_SOYBN
45 69.5 7.3 444 1 PSN_CAEEL

ALIGNMENTS

RESULT 1
CRTZ_PANAN
ID CRTZ_PANAN STANDARD; PRT; 175 AA.
AC P21588
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-carotene hydroxylase.
GN CRTZ.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=20D3;
RX MEDLINE=91072214; PubMed=2254247;
RA Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,
RA Nakamura K., Harashima K.;
RT "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway
RT by functional analysis of gene products expressed in Escherichia
RT coli.";
RL J. Bacteriol. 172:6704-6712(1990).
CC -!- FUNCTION: Catalyzes the hydroxylation reaction from
CC beta-carotene to zeaxanthin.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90087; BAA14129.1; -;
CC PIR; F37802; F37802.
CC InterPro; IPR005596; Carotene_hydrox.
CC InterPro; IPR006087; Sterol_desat.
CC Pfam; PF03897; Carotene_hydrox; 1.
CC Carotenoid biosynthesis.
CC
CC SEQUENCE 175 AA; 19816 MW; 4EB1E57011CS1F9A CRC64;
SQ

Query Match 93.3%; Score 892; DB 1; Length 175;

Best Local Similarity 91.4%; Pred. No. 6.4e-79;

Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLMTWNNALIVFTVVGMEVVAALAHKYIMHGCHGWHLSSHEPRKGAFFVNDLYAVVFAI 60

DB 1 MLMTWNNALIVFTVVGMEVVAALAHKYIMHGCHGWHLSSHEPRKGAFFVNDLYAVVFAA 60

QY 61 VSIALYFGSTGIMPLWIGAGTAYGLLYFMVHDLVHQWPFRYIPRKGYLKLRYMAH 120

DB 61 LSILLIYLGSTGIMPLWIGAGTAYGLLYFMVHDLVHQWPFRYIPRKGYLKLRYMAH 120

QY 121 RMHVAVRGKGCVCVSGFLYAPPLSKQLTLRERHAARSGAARDQDGVDTSSSGK 175

```
|||||
121 RMHVRKGVSGVSGFLYAPPLSKLQTLRHRGARGAARDQCGEDBPASGK 175
:|||||

RESULT 2
RTZ_ERWHE
D_-CRTZ_ERWHE STANDARD; PRT; 176 AA.
C_Q01332;
T_01-APR-1993 (Rel. 25, Created)
T_01-APR-1993 (Rel. 25, Last sequence update)
T_10-OCT-2003 (Rel. 42, Last annotation update)
E_Beta-carotene hydroxylase.
N_CRTZ OR CRTH
S_Erwinia herbicola.
C_Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
X_Enterobacteriaceae; Pantoea.
X_NCBI_TaxID=549;
N_1
P_SEQUENCE FROM N.A.
C_STRAIN=EHO10;
X_MEDLINE=931138098; PubMed=8422926;
A_Hundt B.S., O'Brien D.A., Beyer P., Kleinig H., Hearst J.E.;
T_ "In vitro expression and activity of lycopene cyclase and
T_beta-carotene hydroxylase from Erwinia herbicola.";
L_PBS Lett. 315:329-334(1993).
[2]
P_SEQUENCE FROM N.A.
C_STRAIN=EHO10;
A_Liu S.T.;
L_Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
C_!- FUNCTION: Catalyzes the hydroxylation reaction from
C_beta-carotene to zeaxanthin.
C_!- PATHWAY: Carotenoid biosynthesis.
-----
C_This SWISS-PROT entry is copyright. It is produced through a collaboration
C_between the Swiss Institute of Bioinformatics and the EMBL outstation -
C_the European Bioinformatics Institute. There are no restrictions on its
C_use by non-profit institutions as long as its content is in no way
C_modified and this statement is not removed. Usage by and for commercial
C_entities requires a license agreement (See http://www.isb-sib.ch/announce/
C_or send an email to license@isb-sib.ch).
-----
C_EMBL; M87280; AAA64983.1; -.
C_PIR; S52982; S52982.
C_InterPro; IPR005596; Carotene hydrox.
C_Pfam; PF03897; Carotene hydrox.
C_Carotenoid biosynthesis.
C_SEQUENCE 176 AA; 19966 MW; BDE0A1C8A53F7BA8 CRC64;
Query Match 58.4%; Score 654; DB 1; Length 176;
Best Local Similarity 58.2%; Pred. No. 5.2e-56;
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;
Y_6 NALIVFVTVGVMEVVAALAHKYIMHGHWGSHLSHHEPRKGAPEVNDLYAVVFAIVSIALI 65
b_4 NSLLIVLSVIAEMGIAFTHYIMHGHWGSHLSHHEPRKGFELNDLYAVVFAIVSIALI 63
Y_66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLGVHQRWPFYIPRKGYLRLYVAHRMHA 125
b_64 IAVGTAGWPLQWIGAGMTAYGLLYFMVHDLGVHQRWPFYIPRKGYLRLYVAHRMHA 123
Y_126 VRGKGVSGFLYAPPLSKLQTLRHRGARGAARDQCGEDVDTSSS 173
b_124 VRGKGVSGFLYAPPLSKLQTLRHRGARGAARDQCGEDVDTSSS 173

ESULT 3
RTZ_ALCSP
D_-CRTZ_ALCSP STANDARD; PRT; 162 AA.
C_Q44262;
T_01-NOV-1997 (Rel. 35, Created)
```

```
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-carotene hydroxylase.
GN CRTZ.
OS Alcaligenes sp.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_TaxID=512;
RN 1
P_SEQUENCE FROM N.A.
MEDLINE=95251715; PubMed=7733978;
RA_Misawa N., Kajiura S., Kondo K., Yokoyama A., Satomi Y., Saito T.,
RA_Miki W., Ohtani T.;
RT_ "Canthaxanthin biosynthesis by the conversion of methylene to keto
RT_groups in a hydrocarbon beta-carotene by a single gene.";
RL_Biochem. Biophys. Res. Commun. 209:867-876(1995).
CC_!- FUNCTION: Catalyzes the hydroxylation reaction from beta-carotene
CC_to zeaxanthin via beta-cryptoxanthin (By similarity).
CC_!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
CC_biosynthetic pathway.
-----
C_This SWISS-PROT entry is copyright. It is produced through a collaboration
C_between the Swiss Institute of Bioinformatics and the EMBL outstation -
C_the European Bioinformatics Institute. There are no restrictions on its
C_use by non-profit institutions as long as its content is in no way
C_modified and this statement is not removed. Usage by and for commercial
C_entities requires a license agreement (See http://www.isb-sib.ch/announce/
C_or send an email to license@isb-sib.ch).
-----
C_EMBL; D58422; BAA09597.1; -.
C_InterPro; IPR005596; Carotene hydrox.
C_Pfam; PF03897; Carotene hydrox; 1.
C_Carotenoid biosynthesis.
C_SEQUENCE 162 AA; 19157 MW; OCA5235CBFEB33B5 CRC64;
Query Match 52.2%; Score 499; DB 1; Length 162;
Best Local Similarity 57.8%; Pred. No. 4e-41;
Matches 93; Conservative 22; Mismatches 38; Indels 8; Gaps 4;
Y_8 LIIVFVTVGVMEVVAALAHKYIMHGHWGSHLSHHEPRKGAPEVNDLYAVVFAIVSIALI 66
b_5 LIIVATVAVNELTAYSVHRIMHGHWGSHLSHHEPRKGAPEVNDLYAVVFAIVSIALI 64
Y_67 YFGSTGIWPLQWIGAGMTAYGLLYFMVHDLGVHQRWPFYIPRKGYLRLYVAHRMHA 125
b_65 TVGAYW-WPVLWWIALGWTYGLIYFIHDLGVHQRWPFYIPRKGYLRLYVAHRMHA 123
Y_126 VRGKGVSGFLYAPPLSKLQTLRHRGARGAARDQCGED 166
b_124 VEGDHCVSGFLYAPPLSKLQTLRHRGARGAARDQCGED 159

RESULT 4
CRTZ_AGRAU STANDARD; PRT; 162 AA.
AC_P54973;
DT_01-OCT-1996 (Rel. 34, Created)
DT_01-OCT-1996 (Rel. 34, Last sequence update)
DT_10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-carotene hydroxylase.
GN CRTZ.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=44155;
RN 1
P_SEQUENCE FROM N.A.
MEDLINE=96062243; PubMed=7592436;
RA_Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiura S., Saito T.,
RA_Ohtani T., Miki W.;
RT_ "Structure and functional analysis of a marine bacterial carotenoid
RT_biosynthesis gene cluster and astaxanthin biosynthetic pathway
```

proposed at the gene level.";
J. Bacteriol. 177:6575-6584(1995).
-!- FUNCTION: Catalyzes the hydroxylation reaction from
beta-carotene to zeaxanthin via beta-cryptoxanthin.
-!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
biosynthetic pathway.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; D58420; BRA09592.1; -;
InterPro; IPR005596; Carotene hydros.
InterPro; IPR006087; Sterol desat.
Pfam; PF03897; Carotene hydros; 1.
Carotenoid biosynthesis.
SEQUENCE 162 AA; 18926 MW; F88891A12A92C6C CRC64;

Query Match 51.8%; Score 495; DB 1; Length 162;
Best Local Similarity 58.0%; Pred. No. 9.7e-41;
Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALLIVFTVVCMEVVALAHKYIMHG-WGQGNHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64
3 NFLIVAVGLVMEITATSVSHRWIMHGLPMGKSHHEHDALEKNDLGLYFAIVATV 62

65 LIYVGSTGIW- -LQWIGAGMTAYGLLYFMVHGLVHQRWPPFYIPKGYKLYMAHRM 122
53 LFTVG- -WIRAPVLWALGNTVYGLIYFVLDGLVHQRWPPFYIPKGYARLYQARHL 120

123 HNAVKGEGCVSGFLYAPPLSKLQATLRHRAARSGAARDE 164
121 HHAVEGRDHCVSFGFIYAPPVVDKLQDLK-----MSGVLRAE 157

SULT 5
3E_BRANA

F03E_BRANA STANDARD; PRT; 383 AA.
P48624; P46311;
01-NOV-1995 (Rel. 32, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.19.-).
FAD5.
Brassica napus (Rape).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Brassicales; Brassicaceae; Brassica.
NCBI_TaxID=3708;
[1]
SEQUENCE FROM N.A.
MEDLINE=93088059; PubMed=1455229;
Arondel V., Lemieux B., Hwang I., Gibson S., Goodman H.M.,
Somerville C.R.;
"Map-based cloning of a gene controlling omega-3 fatty acid
desaturation in Arabidopsis.";
Science 258:1353-1355(1992).
[2]
SEQUENCE FROM N.A.
TISSUE=seed;
MEDLINE=94302147; PubMed=8029334;
Yadav N.S., Wierzbicki A., Aegerter M., Caster C.S., Perex-Grau L.,
Kinney A.J., Hitz W.D., Booth J.R. Jr., Schweiger B., Stecca K.L.,
Allen S.M., Blackwell M., Reiter R.S., Carlson T.J., Russell S.H.,
Feldman K.A., Pierce J., Browne J.;
"Cloning of higher plant omega-3 fatty acid desaturases.";
Plant Physiol. 103:467-476(1993).
-!- FUNCTION: ER (microsomal) omega-3 fatty acid desaturase introduces
the third double bond in the biosynthesis of 18:3 fatty acids,

important constituents of plant membranes. It is thought to use
cytochrome b5 as an electron donor and to act on fatty acids
esterified to phosphatidylcholine and, possibly, other
phospholipids.
-!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
-!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
-!- DOMAIN: The histidine box domains may contain the active site
and/or be involved in metal ion binding.
-!- SIMILARITY: Belongs to the fatty acid desaturase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; L01418; AAA32994.1; -;
EMBL; L22962; AAA61775.1; -;
PIR; A44227; A44227.
InterPro; IPR005804; FA desat. fam.
Pfam; PF00487; FA desaturase; 1.
ProDom; PD001081; FA desat fam; 2.
Oxidoreductase; Fatty acid biosynthesis; Endoplasmic reticulum;
Transmembrane.
PT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT DOMAIN 98 102 HISTIDINE BOX-1.
FT DOMAIN 134 138 HISTIDINE BOX-2.
FT DOMAIN 301 305 HISTIDINE BOX-3.
FT DOMAIN 11 11 V -> A (IN REF. 2).
FT CONFLICT 15 22 SGARKERG -> ER (IN REF. 2).
FT CONFLICT 57 57 T -> A (IN REF. 2).
FT CONFLICT 64 64 A -> V (IN REF. 2).
FT CONFLICT 68 68 M -> V (IN REF. 2).
FT CONFLICT 78 78 L -> F (IN REF. 2).
FT CONFLICT 84 84 V -> A (IN REF. 2).
FT CONFLICT 113 114 SV -> TA (IN REF. 2).
FT CONFLICT 162 162 P -> S (IN REF. 2).
FT CONFLICT 181 181 I -> L (IN REF. 2).
FT CONFLICT 195 195 P -> Y (IN REF. 2).
FT CONFLICT 232 232 D -> G (IN REF. 2).
FT CONFLICT 264 264 E -> D (IN REF. 2).
FT CONFLICT 320 321 RA -> KS (IN REF. 2).
SQ SEQUENCE 383 AA; 43936 MW; DAD7C3AG67A12826A CRC64;

Query Match 10.6%; Score 101.5; DB 1; Length 383;
Best Local Similarity 25.5%; Pred. No. 0.018;
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;

QY 44 RKGAPEVNDLYAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHD----- 95
DB 51 RMSYVTRDIFAV--ALAAVAVFDWFLNPLVYVAQG-TLFWAIFVLGHDCGHGSFSD 107

QY 96 -----GLVHQRWPPFYIPKGYKLYMAHRMHAVRG 128
DB 108 IPLLNSVVGHILHS---FILVPYHGW----RISHRTTHQNHG 142

RESULT 6
F03E_ARATH
ID F03E_ARATH STANDARD; PRT; 386 AA.
AC P48623;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.19.-).
GN FAD3 OR ARZG2990 OR F23F1.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

C eurosids II; Brassicales; Brassicaceae; Arabidopsis.
X NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=cv. Columbia; TISSUE=Seedling;
X MEDLINE=94302147; PubMed=8029334;
A Yadav N.S., Wierzbicki A., Aegerter M., Perez-Grau L.,
A Kinnay A.J., Hitz W.D., Booth J.R. Jr., Schweiger B., Stecca K.L.,
A Allen S.M., Blackwell M., Ratter R.S., Carlson T.J., Russell S.H.,
A Feldmann K.A., Pierce J., Browe J.,
T "Cloning of higher plant omega-3 fatty acid desaturases."
L Plant Physiol. 103:467-476(1993).
N [2]
P SEQUENCE FROM N.A.
C STRAIN=cv. Columbia; TISSUE=Hypocotyl;
X Watanabe M.C., Yamamoto K.T.;
L Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
N [3]
P SEQUENCE FROM N.A.
C STRAIN=cv. Columbia;
X MEDLINE=94345020; PubMed=8066143;
A Nishibuchi T., Nishimura M., Aronoff V., Iba K.;
T "Genomic nucleotide sequence of a gene encoding a microsomal omega-3
L fatty acid desaturase from Arabidopsis thaliana."
L Plant Physiol. 105:767-768(1994).
N [4]
P SEQUENCE FROM N.A.
C STRAIN=cv. Columbia;
X MEDLINE=20083487; PubMed=10617197;
A Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
A Fujii C.F., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
A Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
A Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayah L.,
A Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Cressy T.H.,
A Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
A Nierman W.C., White O., Zisen J.A., Salzberg S.L., Fraser C.M.,
A Venter J.C.;
T "Sequence and analysis of chromosome 2 of the plant Arabidopsis
L thaliana."
L Nature 402:761-768(1999).
N [5]
P SEQUENCE FROM N.A.
C STRAIN=cv. Columbia;
X MEDLINE=22954850; PubMed=14593172;
A Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
A Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.P.,
A Karlins-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
A Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
A Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
A Arakawa T., Ban J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
A Chao Q., Choy N., Anju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
A Hayashizaki Y., Johnson-Ropson C., Hsuan V.W., Iida K., Karnes M.,
A Khan S., Koeseema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
A Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
A Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
A Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
T "Empirical analysis of transcriptional activity in the Arabidopsis
L genome."
L Science 302:842-846(2003).
X -!- FUNCTION: Microsomal (ER) omega-3 fatty acid desaturase introduces
the third double bond in the biosynthesis of 18:3 fatty acids,
important constituents of plant membranes. It is thought to use
cytochrome b5 as an electron donor and to act on fatty acids
esterified to phosphatidylcholine and, possibly, other
phospholipids.
X -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
X -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
X -!- ALTERNATIVE PRODUCTS:
Event-Alternative splicing; Named isoforms=1;
Comment=A number of isoforms are produced. According to EST
sequences;
X Name=1;
X IsoId=P48623-1; Sequence=Displayed;

CC -!- TISSUE SPECIFICITY: Abundant in leaves and seedlings. Barely
CC detectable in root tissue.
CC -!- DOMAIN: The histidine box domains may contain the active site
CC and/or be involved in metal ion binding.
CC -!- SIMILARITY: Belongs to the fatty acid desaturase family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L22931; AAA61778.1; -
CC EMBL: D17579; BAA04505.1; -
CC EMBL: D26508; BAA05514.1; -
CC EMBL: AC004680; AAC31854.1; -
CC EMBL: AY063866; AAL36322.1; -
CC EMBL: AY096462; AAM20102.1; -
CC PIR: JQ2335; JQ2335.
CC InterPro: IPR005804; FA_desat_fam.
CC Pfam: PF00487; FA_desaturase; 1.
CC ProDom: PD001081; FA_desat_fam; 2.
X Fatty acid biosynthesis; Oxidoreductase; Endoplasmic reticulum;
X Transmembrane; Alternative splicing.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT DOMAIN 101 105 HISTIDINE BOX-1.
FT DOMAIN 137 141 HISTIDINE BOX-2.
FT DOMAIN 304 308 HISTIDINE BOX-3.
SQ SEQUENCE 386 AA; 44076 MW; 6A7EA2A692885164 CRC64;
Query Match 9.8%; Score 93.5; DB 1; Length 386;
Best local Similarity 26.5%; Pred. NO. 0.1;
Matches 27; Conservative 17; Mismatches 31; Indels 27; Gaps 5;
Qy 44 RKCAFENVLYAVFAIVSTALTYFGSTGIMPLQWIGAGMTAYGLLYPMVHD----- 95
Db 54 RSMYVVRDIIAV--AALAAVYVDVSWFENLYWAAQG-TLFWAIFVLGCHDCGHGSPD 110
Qy 96 -----GLVHQRPFPYIPRKGLKLYMAHRMHAVRG 128
Db 111 IPLLNSVVGHLHS---FILVPYHGW----RISHRTHQNHG 145
RESULT 7
CAP MYCLE STANDARD; PRT; 934 AA.
ID P46710; Q9CCN5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC).
GN PPC OR MLO578 OR B1496_C3_207.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769,
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RC MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeter K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,

01-AUG-1992 (Rel. 23, Created)
 01-AUG-1992 (Rel. 23, Last sequence update)
 10-OCT-2003 (Rel. 42, Last annotation update)
 Longevity assurance homolog 1 (UOG-1 protein) (LAG1 protein).
 LASSI OR UOG1 OR LAG1.
 Homo sapiens (human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A. (ISOFORM 1).
 MEDLINE=91239545; PubMed=2034669;
 Lee S.-J.;
 "Expression of growth/differentiation factor 1 in the nervous system:
 conservation of a bicistronic structure.";
 Proc. Natl. Acad. Sci. U.S.A. 88:4250-4254 (1991).
 [2]
 SEQUENCE FROM N.A. (ISOFORM 1).
 MEDLINE=93091901; PubMed=9872981;
 Jiang J.C., Kirchman P.A., Zgulski M., Hunt J., Jaswinski S.M.;
 "Homologs of the yeast longevity gene LAG1 in *Caenorhabditis elegans*
 and human.";
 Genome Res. 8:1259-1272 (1998).
 [3]
 SEQUENCE FROM N.A.
 Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
 Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
 Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,
 Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 Kronmiller B., Arellano A., Montgomerly M., Ow D., Nolan M., Trong S.,
 Kobayashi A., Olsen A.S., Carrano A.V.;
 Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A. (ISOFORM 2).
 TISSUE=Hypothalamus;
 MEDLINE=22388257; PubMed=12477932;
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.N., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 Schnerch A., Schein J.E., Jones S.J.N., Marra M.A.;
 "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1;
 IsoId=27544-1; Sequence=Displayed;
 Name=2;
 IsoId=27544-2; Sequence=VSP_003049;
 Note=No experimental confirmation available;
 -!- SIMILARITY: Contains 1 TFC (TRAM/LAG1/CLN8) domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

 DR EMBL; M62302; AAA58500.1; -;
 DR EMBL; AF105009; AAD16892.1; -;
 DR EMBL; AF105005; AAD16892.1; JOINED.
 DR EMBL; AF105006; AAD16892.1; JOINED.
 DR EMBL; AF105007; AAD16892.1; JOINED.
 DR EMBL; AF105008; AAD16892.1; JOINED.
 DR EMBL; AF105009; AAD16892.1; JOINED.
 DR EMBL; AC005197; AAC24611.1; -;
 DR EMBL; AC003972; -; NOT ANNOTATED_CDS.
 DR EMBL; BC022450; AAH22450.1; -;
 DR PIR; D39364; D39364.
 DR Genew; HGNC:14253; LASS1.
 DR MIM; 606919; -;
 DR GO; GO:0016021; C: integral to membrane; ISS.
 DR GO; GO:0007568; P: aging; ISS.
 DR InterPro; IPR005547; LAG1.
 DR InterPro; IPR006634; TFC.
 DR Pfam; PF03798; LAG1; 1.
 DR SMART; SM00724; TFC; 1.
 DR PROSITE; PS00922; TFC; 1.
 KW Transmembrane; Alternative splicing.
 FT TRANSMEM 53 73 POTENTIAL.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 176 196 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT TRANSMEM 287 307 POTENTIAL.
 FT DOMAIN 97 311 TFC.
 FT VARSPLIC 338 350 Missing (in isoform 2).
 FT CONFLICT 111 111 /FTid=VSP_003049.
 FT SEQUENCE 350 AA; 39536 MW; F102C12C47DB4162 CRC64;
 SQ
 Query Match 8.7%; Score 83; DB 1; Length 350;
 Best Local Similarity 22.2%; Pred. No. 0.97;
 Matches 42; Conservative 18; Mismatches 53; Indels 76; Gaps 8;
 QY 3 WENALIVFVT-----VVGMEVVALAHKYMEGWGW-----W 36
 DB 113 WSYAYLLFGDYDFHDPFVFDVTFGMAVPRDIAAYLLQGSFYGHSHIYATLYMDTW 172
 QY 37 H-----LSHHEPRKGAFEVNDLVAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLY 90
 DB 173 REDSWMLLH-----VVTLLIVSSVAFRYEN-----VGILV 205
 QY 91 FMVHD-----GLVHOMFPRIPKGYLKRLYMAHMEHVAHVGEGCVSPF-----LYA 140
 DB 206 LFLHDSIDVQLFETKLIYFKSRGGSYHRL-----HALAADIGCLSGFSGFWFELYW 258
 QY 141 PPLSKLOAT 149
 DB 259 PPLKVLAT 267
 RESULT 12
 PD3C ARATH
 ID PD3C ARATH STANDARD; PRT; 446 AA.
 AC P46310;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-).
 QS FAD7 OR FADD OR AT3G1170 OR F9F8.4 OR F11B9.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Hypocotyl;

X MEDLINE=94302147; PubMed=8029334;
 A Yadav N.S., Wierzbicki A., Aegerter M., Caster C.S., Perez-Grau L.,
 A Kinney A.J., Hitz J.D., Booth R.R. Jr., Schweiger B., Stecca K.L.,
 A Allen S.M., Blackwell M., Reiter T.J., Carlson T.J., Russell S.H.,
 A Feldmann K.A., Pierce J., Browne J.,
 T "Cloning of higher plant omega-3 fatty acid desaturases";
 L Plant Physiol. 103:467-476(1993).
 N [2]
 P SEQUENCE FROM N.A.
 C STRAIN=cv. Columbia; TISSUE=Aerial parts;
 X MEDLINE=94043239; PubMed=8226956;
 A Iba K., Gibson S., Nishiuchi T., Fuse T., Nishimura M., Arondel V.,
 A Hugly S., Somerville C.R.;
 T "A gene encoding a chloroplast omega-3 fatty acid desaturase
 I complements alterations in fatty acid desaturation and chloroplast
 I copy number of the fat7 mutant of Arabidopsis thaliana";
 N J. Biol. Chem. 268:24099-24105(1993).
 N [3]
 P SEQUENCE FROM N.A.
 C STRAIN=cv. Columbia; TISSUE=Hypocotyl;
 X MEDLINE=94043239; PubMed=8226956;
 A Watabiki M., Yamamoto K.;
 L Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 N [4]
 P SEQUENCE FROM N.A.
 C STRAIN=cv. Columbia;
 X MEDLINE=21016720; PubMed=11130713;
 A Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,
 A Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 A Dalseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 A De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
 A Wincker P., Catolico L., Weissbach J., Saurin W., Quetier F.,
 A Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 A Wurmbach E., Drzonek H., Erfle H., Jordan N., Brandt S.,
 A Wiedelmann R., Kranz H., Voss H., Holland N., Bangert S.,
 A Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 A Conrad A., Hornischer K., Kauer G., Loehert T.-H., Nordleik G.,
 A Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
 A Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 A Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy D.,
 A De Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 A Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
 A Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 A Mayer K.F.X., Kaul S., Town C.D., Koo H.E., Tallon L.J., Jenkins J.,
 A Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 A Creasy T.H., Haas B., Maiti A., Wu D., Peterson J., Van Aken S.,
 A Pal G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 A Pruss D., Lin X., Niemman W.C., Salzberg S.L., White O., Venter J.C.,
 A Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 A Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 A Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,
 A Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
 A Watanabe A., Yamada M., Yasuda M., Tabata S.;
 T "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 I thaliana";
 IL Nature 408:820-822(2000).
 X -!- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces
 X the third double bond in the biosynthesis of 16:3 and 18:3 fatty
 X acids, important constituents of plant membranes. It is thought
 X to use ferredoxin as an electron donor and to act on fatty acids
 X esterified to galactolipids, sulfolipids and phosphatidylglycerol.
 X -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
 X -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
 X -!- TISSUE SPECIFICITY: Most abundant in leaves and seedlings.
 X -!- DOMAIN: The histidine box domains may contain the active site
 X and/or be involved in metal ion binding.
 X -!- SIMILARITY: Belongs to the fatty acid desaturase family.
 X
 X This SWISS-PROT entry is copyright. It is produced through a collaboration
 X between the Swiss Institute of Bioinformatics and the EMBL outstation -
 X the European Bioinformatics Institute. There are no restrictions on its
 X use by non-profit institutions as long as its content is in no way
 X modified and this statement is not removed. Usage by and for commercial
 X entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 DR ENBL; L22961; AAA61773.1; -
 DR ENBL; D14007; BAA03106.1; -
 DR ENBL; D26019; BAA05040.1; -
 DR ENBL; AC009991; AAF01508.1; -
 DR ENBL; AC073395; AAG50977.1; -
 DR PIR; JQ2336; JQ2336.
 DR InterPro: IPR005804; FA desat fam.
 DR Pfam: PF00487; FA desaturase; 1.
 DR ProDom: PD001091; FA desat fam; 2.
 KM Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
 KM Transit peptide.
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT CHAIN 1 446 OMEGA-3 FATTY ACID DESATURASE.
 FT DOMAIN 163 167 HISTIDINE BOX-1.
 FT DOMAIN 199 203 HISTIDINE BOX-2.
 FT DOMAIN 366 370 HISTIDINE BOX-3.
 SQ SEQUENCE 446 AA; 51174 MW; 121125F634553D35 CRC64;
 Query Match 8.6%; Score 82.5; DB 1; Length 446;
 Best Local Similarity 23.3%; Pred. No. 1.4;
 Matches 30; Conservative 24; Mismatches 36; Indels 39; Gaps 8;
 Qy 18 EVYVAL-AHKYIMGHWGWHLSHHEPRKGAFEVNDLVAVFAIVSIALIVFGSTGWPL 76
 Db 100 DIFAAIPKHCWKNPW-----KSLSYVVRDV-AIVFALAAGA-AYLNWVWPL 146
 Qy 77 QMIGAGTAYGLLYFMVHD-----GLVHQRPFRYPYPRKGYLRLYMA 119
 Db 147 YMLAQG-TMFWALFVLGHGCHGSFNSDKLSNVVGHLLRSS---ILVPHGW----RIS 198
 Qy 120 HRVHVAVRG 128
 Db 199 HRTTHQMHG 207
 RESULT 13
 CYOB_BUCAI STANDARD; PRT; 562 AA.
 ID P57543;
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ubiquinol oxidase polypeptide I (EC 1.10.3.-) (Cytochrome O subunit 1)
 DE (Oxidase BO(3) subunit 1) (Cytochrome O ubiquinol oxidase subunit 1).
 DE CYOB OR BU471.
 GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. APS";
 RL Nature 407:81-85(2000).
 CC -!- FUNCTION: Cytochrome O terminal oxidase complex is the component
 CC of the aerobic respiratory chain that predominates when cells are
 CC grown at high aeration. This ubiquinol oxidase shows proton pump
 CC activity across the membrane in addition to the electron transfer
 CC (By similarity).
 CC -!- CATALYTIC ACTIVITY: Ubiquinol-8 + O(2) = Ubiquinone-8 + H(2)O.
 CC -!- COFACTOR: Contains two protoheme IX (heme B55 and B562) and copper
 CC B (By similarity).
 CC -!- PATHWAY: Ubiquinol oxidase catalyzes the terminal step in the
 CC electron transport chain.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AP001119; BAB13168.1; -
HSSP; P18401; 1PFT.
InterPro; IPR000983; COX1.
Pfam; PF00115; COX1.1.
PRINTS; PR01165; CYCOXIDASI.
PROSITE; PS00077; COX1.1.
Oxidoreductase; Respiratory chain; Heme; Copper;
Hydrogen ion transport; Complete;
DOMAIN 1 14
TRANSMEM 15 35
EXTRACELLULAR (POTENTIAL).
DOMAIN 36 58
CYTOPLASMIC (POTENTIAL).
DOMAIN 59 79
EXTRACELLULAR (POTENTIAL).
TRANSMEM 107 127
EXTRACELLULAR (POTENTIAL).
DOMAIN 128 145
CYTOPLASMIC (POTENTIAL).
TRANSMEM 146 166
EXTRACELLULAR (POTENTIAL).
DOMAIN 167 189
EXTRACELLULAR (POTENTIAL).
TRANSMEM 190 210
EXTRACELLULAR (POTENTIAL).
DOMAIN 211 232
CYTOPLASMIC (POTENTIAL).
TRANSMEM 233 253
EXTRACELLULAR (POTENTIAL).
DOMAIN 254 277
EXTRACELLULAR (POTENTIAL).
TRANSMEM 278 298
EXTRACELLULAR (POTENTIAL).
DOMAIN 299 309
CYTOPLASMIC (POTENTIAL).
TRANSMEM 310 330
EXTRACELLULAR (POTENTIAL).
DOMAIN 331 346
EXTRACELLULAR (POTENTIAL).
TRANSMEM 347 367
EXTRACELLULAR (POTENTIAL).
DOMAIN 368 380
CYTOPLASMIC (POTENTIAL).
TRANSMEM 381 401
EXTRACELLULAR (POTENTIAL).
DOMAIN 402 413
EXTRACELLULAR (POTENTIAL).
TRANSMEM 414 434
EXTRACELLULAR (POTENTIAL).
DOMAIN 435 456
EXTRACELLULAR (POTENTIAL).
TRANSMEM 457 477
EXTRACELLULAR (POTENTIAL).
DOMAIN 478 493
EXTRACELLULAR (POTENTIAL).
TRANSMEM 494 514
EXTRACELLULAR (POTENTIAL).
DOMAIN 515 586
CYTOPLASMIC (POTENTIAL).
TRANSMEM 587 607
EXTRACELLULAR (POTENTIAL).
DOMAIN 608 608
EXTRACELLULAR (POTENTIAL).
TRANSMEM 609 629
EXTRACELLULAR (POTENTIAL).
DOMAIN 630 662
CYTOPLASMIC (POTENTIAL).
METAL 106 106
IRON (HEME B AXIAL LIGAND) (PROBABLE).
METAL 284 284
COPPER B (PROBABLE).
METAL 288 288
COPPER B (PROBABLE).
METAL 333 333
COPPER B (PROBABLE).
METAL 334 334
COPPER B (PROBABLE).
METAL 419 419
IRON (HEME O AXIAL LIGAND) (PROBABLE).
METAL 421 421
IRON (HEME B AXIAL LIGAND) (PROBABLE).
1'-histidyl-3'-tyrosine (His-Tyr
(by similarity).
CROSSLNK 284 288
SEQUENCE 662 AA; 75455 MW; 3F502A628133EA65 CRC64;

Query Match 8.6%; Score 82; DB 1; Length 662;
Best Local Similarity 20.4%; Pred. No. 2.3;
Matches 39; Conservative 25; Mismatches 53; Indels 74; Gaps 12;
1 MLTWNALIVFTV-----VGMVEVAALAHK-----YTMHNG-----WGNHLSH 40
278 LITWGHDEVILVLPVGVPSVNVATFSKRLPGYSLVNATILSFIWV-----LH 333
41 HEPKRGK-PEVNDLYAVVPAVSIALYFGSTGWPLOWIGAGTAYGLLYFMVHDLVH 99
334 HFTVAGADVNTFFGTTMTIAL-----PTGVKIFNWL-----FTIYQGRVH 376
100 QR-----WPFY-----IPRKYL--KELYMAHMHAVRGKE--GCVS-- 134
377 MESSILWLGFLVTFPSIGMTGVLVSPVPPADFLVHNSLFLVAHFHNVILGGVVGCFAGI 436

QY 135 -----RGFL 138
DB 437 NYWFFKLFSGV 447
RESULT 14
FD3C RICCO STANDARD; PRT; 460 AA.
ID PD3C RICCO
AC P48619;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-).
GN PAD7A-1.
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosids;
OC Euphorbiales; Malvaceae; Euphorbiaceae; Acalyphoideae; Acalyphaceae;
OC Ricinus.
OX NCBI_TaxID=3998;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Baker 296; TISSUE=Seed;
RX MEDLINE=94302177; PubMed=8029360;
RA van de Ioo P.J.; Somerville C.R.;
RT "Plasmid omega-3 fatty acid desaturase cDNA from Ricinus communis.";
RL Plant Physiol. 195; 443-444 (1994).
CC -!- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces the third double bond in the biosynthesis of 16:3 and 18:3 fatty acids, important constituents of plant membranes. It is thought to use ferredoxin as an electron donor and to act on fatty acids esterified to galactolipids, sulfolipids and phosphatidylglycerol.
CC -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC -!- DOMAIN: The histidine box domains may contain the active site and/or be involved in metal ion binding.
CC -!- SIMILARITY: Belongs to the fatty acid desaturase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L25897; AAA73511.1; -
PIR; T10063; T10063.
InterPro; IPR005804; FA desat fam.
DR Pfam; PF00487; FA desaturase; 1.
DR ProDom; PD001081; FA desat fam; 2.
KW Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 460 OMEGA-3 FATTY ACID DESATURASE.
FT DOMAIN 177 181 HISTIDINE BOX-1.
FT DOMAIN 213 217 HISTIDINE BOX-2.
FT DOMAIN 380 384 HISTIDINE BOX-3.
SQ SEQUENCE 460 AA; 52561 MW; 836592904E93C7B0 CRC64;
Query Match 8.5%; Score 81.5; DB 1; Length 460;
Best Local Similarity 23.3%; Pred. No. 1.8;
Matches 30; Conservative 22; Mismatches 38; Indels 39; Gaps 8;
QY 18 EVVAAL-AHKYIMHGNGWHLSSHHEPKGAFVNDLYAVVPAVSIALYFGSTGWL 76
DB 114 DIRAIPKHCHVQNPW-----RSMYSVLRDV-VVFGLAAVA-AYFNWVAVPL 160
QY 77 QWIGAGTAYGLLYFMVH-----GLVHQRWPFYIPKGYLKRLVMA 119
DB 161 YWFCQG-TMFWALFVLGHDCHGSGFSNNPKLNSVVGHLHSS---ILVPYHG---RIS 212

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

% protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 4.87104 Seconds
(without alignments)
3455.835 Million cell updates/sec

title: US-09-941-947A-36

effect score: 956

sequence: 1 MLWLNALIVFTVVGMEV.....ARSGARDEQGVDTSSGX 175

scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

sarched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	892	93.3	175	2 F37802	criz protein - Erw
2	654	68.4	176	2 S52982	beta-carotene hydr
3	271	28.3	151	2 H30469	beta carotene hydr
4	253	26.5	310	2 T09562	beta-carotene hydr
5	101.5	10.6	383	1 A44227	omega-3 fatty acid
6	99.5	10.4	377	1 J02337	omega-3 fatty acid
7	98.5	10.3	410	2 S75695	hypothetical prote
8	93.5	9.8	386	1 J02335	omega-3 fatty acid
9	93	9.7	418	2 J07872	stearyl-CoA 9-des
10	91	9.5	304	2 A26215	conserved hypotet
11	91	9.5	313	2 E97397	hypothetical prote
12	88	9.2	301	2 G83556	hypothetical prote
13	87	9.1	907	2 S72765	phosphoenolpyruvat
14	87	9.1	934	2 B86981	probable phosphoen
15	84.5	8.8	162	2 A3586	hypothetical membr
16	84.5	8.8	404	2 P00812	omega-3 fatty acid
17	84.5	8.8	455	2 T44519	putrescine/ornithi
18	83.5	8.7	172	2 T11298	NADH2 dehydrogenas
19	83	8.7	350	2 D39364	GDF-1 embryonic gr
20	82.5	8.6	441	2 T03029	omega-3 fatty acid
21	82.5	8.6	446	1 J02336	omega-3 fatty acid
22	82	8.6	662	2 H84984	bo-type ubiquinol
23	81.5	8.5	460	2 T10063	omega-3 fatty acid
24	79.5	8.3	567	2 A83023	potassium-transpor
25	79.5	8.3	567	2 H38261	hypothetical prote
26	78.5	8.2	431	2 T07685	omega-3 fatty acid
27	78	8.2	415	2 AG2044	hypothetical prote
28	78	8.2	551	2 A87019	probable cytochrom
29	78	8.2	751	2 A81816	nitric-oxide reduc

30 78 8.2 751 2 D81062 nitric oxide reduc
31 77.5 8.1 228 2 AB3510 molybdenum transpo
32 77.5 8.1 316 2 T19435 hypothetical prote
33 77.5 8.1 410 2 A70648 Probable NADH2 deh
34 77.5 8.1 489 2 E84229 DNA damage-inducib
35 77 8.1 222 1 D69991 conserved hypotet
36 77 8.1 3473 1 A46112 genome polyprotein
37 77 8.1 32927 2 S27927 polyprotein - rice
38 76.5 8.0 332 1 Q0BE39 BGLF3 protein - hu
39 76.5 8.0 345 2 S72490 N-acetyl-gamma-glu
40 76.5 8.0 453 1 J02339 omega-3 fatty acid
41 76.5 8.0 651 2 S44257 phosphotransferase
42 76 7.9 156 2 S08427 12K protein mlai2A
43 76 7.9 208 2 T23328 hypothetical prote
44 76 7.9 300 2 H87631 integral membrane
45 76 7.9 542 2 C70732 probable integral

ALIGNMENTS

RESULT 1

F37802

criz protein - Erwinia uredovora

C:Species: Erwinia uredovora

C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000

C:Accession: F37802

R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim

J. Bacteriol. 172, 6704-6712, 1990

A:Title: Clucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio

A:Reference number: A37802; MUID:91072214; PMID:2254247

A:Accession: F37802

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-175 <MIS>

A:Cross-references: GB:D90087; MID:g216681; PIDN:BAAL4129.1; PID:g216687

Query Match 93.3%; Score 892; DB 2; Length 175;

Best Local Similarity 91.4%; Pred No. 1,8e-81;

Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MLWLNALIVFTVVGMEVVAALAHKYMKGNGWGLSHHEPRKGAPEVNDLVAVPAL 60

Db 1 MLWLNALIVFTVVGMEVVAALAHKYMKGNGWGLSHHEPRKGAPEVNDLVAVPAA 60

Qy 61 VSIALTYFGSTGIMPLQWTGAGMTAYGLLYFMVHDLGVHQRWPERYIPKGYLKLVMH 120

Db 61 LSILLIYLGSTGIMPLQWTGAGMTAYGLLYFMVHDLGVHQRWPERYIPKGYLKLVMH 120

Qy 121 RMHHAVRKEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEQGVDTSSSGX 175

Db 121 RMHHAVRKEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEQGVDTSSSGX 175

RESULT 2

S52982

beta-carotene hydroxylase - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 08-Oct-1999

R:Hundle, B.; Alberti, M.; Nivelsstein, V.; Beyer, P.; Kleinig, H.; Armstrong, G.A.; Bur

Mol. Gen. Genet. 245, 406-416, 1994

A:Title: Functional assignment of Erwinia herbicola Eho 10 carotenoid genes expressed in

A:Reference number: S52976; MUID:95107236; PMID:7808389

A:Accession: S52982

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <HUN>

A:Cross-references: EMBL:M87280; NID:gl48404; PIDN:AAA64983.1; PID:gl48414

Query Match 68.4%; Score 654; DB 2; Length 176;

Best Local Similarity 68.2%; Pred No. 9e-58;

Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

6 NALIVFTVVGMEVVAALAHKYYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAIVSTAL 65
147 VGMFWARWAHRAHSHASLWNNHSHKREGEPPFELNDVFAIVNAPGPAIGLLSYGFENK 206
73 IWPQLQWAGM--TAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAHRMHAVRGKE 130
207 IVPGLCFGAGLITVPGIAYMFVHDGLVHKRPVGPFIADVPYLRKVAHAHQHHT--DKF 264
131 GCVSFGFLYAP 141
265 NGVPYGLFLGP 275

RESULT 5
A44227
omega-3 fatty acid desaturase (EC 1.14.99.-) [similarity] - rape
N;Alternate names: omega-3 linoleate desaturase
C;Species: Brassica napus (rape)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: A44227
R;Arandel V.; Lemieux, B.; Hwang, I.; Gibson, S.; Goodman, H.M.; Somerville, C.R.
Science 258, 1353-1355, 1992
A;Title: Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis
A;Reference number: A44227; PMID:93088059; PMID:1455229
A;Accession: A44227
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-383 <ARO>
A;Cross-references: GB:101418; NID:gl67147; PIDN:AAA32994.1; PID:gl67148
A;Experimental source: developing seed
A;Note: sequence extracted from NCBI backbone (NCBIP:119842)
C;Superfamily: omega-3 fatty acid desaturase
C;Keywords: oxidoreductase

Query Match 10.6%; Score 101.5; DB 1; Length 383;
Best Local Similarity 25.5%; Pred. NO. 0.021;
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;

44 RKGAPEVNDLYAVVFAIVSIALIYFGSTGIMPLQWIGAGMTAYGLLYFMVHD----- 95
51 RMSVYTRDIFAV--AALAMAAVYFDSWFLNPLYVVAQG--TLFNAIFVLGHDCGSGHSPSD 107
96 -----GLVQRPFRYIPRKGYLKRLYMAHRMHAVRG 128
108 IPLNSVVGHLHS---FILVPYHGM---RISHRTHQNHG 142

RESULT 6
JQ2337
omega-3 fatty acid desaturase (EC 1.14.99.-) BN3 [similarity] - rape
C;Species: Brassica napus (rape)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: JQ2337
R;Yadav, N.S.; Wierzbicki, A.; Aegerter, M.; Caster, C.S.; Perez-Grau, L.; Kinney, A.J.;
J.; Russell, S.H.; Feldmann, K.A.; Pierce, J.; Browne, J.
Plant Physiol. 103, 467-476, 1993
A;Title: Cloning of higher plant omega-3 fatty acid desaturases.
A;Reference number: JQ2335; PMID:94302147; PMID:8029334
A;Contents: cDNA:BN3
A;Accession: JQ2337
A;Molecule type: mRNA
A;Residues: 1-377 <YAD>
A;Cross-references: GB:L22962; NID:G408491; PIDN:AAA61775.1; PID:G408492
C;Comment: This enzyme introduces the third double bond in the biosynthesis of 18:2 and
C;Superfamily: omega-3 fatty acid desaturase
C;Keywords: oxidoreductase

Query Match 10.4%; Score 99.5; DB 1; Length 377;
Best Local Similarity 26.3%; Pred. NO. 0.032;
Matches 26; Conservative 17; Mismatches 35; Indels 21; Gaps 4;

44 RKGAPEVNDLYAVVFAIVSIALIYFGSTGIMPLQWIGAGMTAYGLLYFMVHD----- 95

6 NALIVFTVVGMEVVAALAHKYYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAIVSTAL 65
4 NSLIVLSVIMBGIAPTHRYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAIVSTAL 63
66 IYFGSTGIMPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAHRMHAVRGKE 125
64 IAVGTAGVPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAHRMHAVRGKE 123
126 VRKGCVSFGFLYAPFLSKLQATLERRHA--ARSGAARDEQDGVDTSS 173
124 VRKGCVSFGFLYAPFLSKLQATLERRHA--ARSGAARDEQDGVDTSS 173

RESULT 3
90469
eta-carotene hydroxylase (crt2) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: H90469
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
J.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
unpublished to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90469
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-151 <KUR>
A;Cross-references: GB:AE006641; NID:g13816278; PIDN:AAK43015.1; GSPDB:GN00155
C;Genetics:
C;Gene: crt2

Query Match 28.3%; Score 271; DB 2; Length 151;
Best Local Similarity 38.8%; Pred. NO. 1e-19;
Matches 56; Conservative 26; Mismatches 49; Indels 14; Gaps 3;

1 MLMTWALIVFTVVGMEVVAALAHKYYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60
1 MMLIYVGMVAVLTVGVMEFVARLAKHYVHGLWFLTHEDHKEKQAELEKNDLGLVFA 60
61 VSIALYFGSTGIMPLQWIGAGMTAYGLLYFMVHDGLVHOR-----WPFYIPRKGYLK 114
61 VSVYLFPLGLIGQSYVALSIAIGSSYGIAYFFTHDMVTHDRLHLRLSMGLGRP----FK 116
115 RLYMAHRMHAVRGKCVSFGFLY 139
117 DLILVHDHDK---KEGKGNWGLFLP 137

RESULT 4
T09562
eta-carotene hydroxylase homolog L73G19.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T09562
R;Bavan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.P.X.; Schuel
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16098
A;Accession: T09562
A;Molecule type: DNA
A;Residues: 1-310 <RAA>
A;Cross-references: EMBL:AL050400; GSPDB:GN00062; ATSP:L73G19.80
A;Experimental source: cultivar Columbia; EAC clone L73G19
C;Genetics:
C;Gene: ATSP:L73G19.80
C;Map position: 4
C;Introns: 126/3; 146/3; 215/3; 257/3; 276/3

Query Match 26.58%; Score 253; DB 2; Length 310;
Best Local Similarity 41.2%; Pred. NO. 1.4e-17;
Matches 54; Conservative 24; Mismatches 47; Indels 6; Gaps 3;

45 RMSVVDIFAVV--ALAVAVYEDSFFWFLYAAQG-TLFWAIFVLGHDCGHSFSD 101
96 -----GLVHQRPYPRYPRKGYLKLRYMAHRMHHAVRG 128
102 IPLINTAVGHILHSFILVPYHGW-----RISHRTHQNHG 136

RESULT 7
15695
Protophical protein all1376 - Synochocystis sp. (strain PCC 6803)
Species: Synochocystis sp.
Variety: PCC 6803
Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
Accession: S75695
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
A Res. 3, 109-136, 1996
Title: Sequence analysis of the genome of the unicellular cyanobacterium Synochocystis
Reference number: S74322; MUID:97061201; PMID:8905231
Accession: S75695
Status: nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-410 <RAN>
Cross-references: EMBL:D90912; GB:AB001339; NID:gl653228; PIDN:BAA18256.1; PID:dl01898
Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 10.3%; Score 98.5; DB 2; Length 410;
Best Local Similarity 22.6%; Pred. No. 0.044;
Matches 51; Conservative 27; Mismatches 51; Indels 97; Gaps 14;

1 MIWNNALIVFVVGVMEVA-----ALAH-----XYIMHGWGNGWHLHHBPRKGFV-- 50
||| : : : : : ||| : : : : :
3 MAWLMGLGLALASVLWELVDRDCVHALARWSPLYRLKGM-----HHRVFRSDLSWS 55
||| : : : : : ||| : : : : :
51 NDLYA-----VWPAIVSIALIYFGSTGWP--LWIGAGMTAY-----CLLY-- 90
||| : : : : : ||| : : : : :
56 TEIYQKAWYNDVPERLWAF-----GIWPPPLTWVWQFSQWPLILASAGWYILG 109
||| : : : : : ||| : : : : :
91 FMVH-----DGLVHQRPYPRYPRKGYLKLRYMAHRMHH----- 124
||| : : : : : ||| : : : : :
110 FLISALARGVGLPNADITDLTRPGPLTPPAFWVWVRY--HWRHFDPPNAYFGTL 167
||| : : : : : ||| : : : : :
125 -----AVRKGECVGS-----FGLVAPPLSKLQATLRPH 154
||| : : : : : ||| : : : : :
168 TLVDXNLGTLALSGLKXIAVTGAGGCGF-----QALLQLHL 203
||| : : : : : ||| : : : : :

RESULT 8
22335
Omega-3 fatty acid desaturase (EC 1.14.99.-) CF3 [similarity] - Arabidopsis thaliana
Alternate names: protein F23F1.10
Species: Arabidopsis thaliana (mouse-ear cress)
Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Feb-2001
Accession: JQ2335; JQ2487; A84703
Vadav, N.S.; Wierzbicki, A.; Aegerter, M.; Caster, C.S.; Perez-Grau, L.; Kinney, A.J.;
Russell, S.H.; Feldmann, K.A.; Pierce, J.; Browne, J.
Ant. Physiol. 103, 467-476, 1993
Title: Cloning of higher plant omega-3 fatty acid desaturases.
Reference number: JQ2335; MUID:94302147; PMID:8029334
Accession: JQ2335
Molecule type: mRNA
Residues: 1-386 <YAD>
Cross-references: GB:D17579; NID:91030693; PIDN:BAA04505.1; PID:q471091
Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
Submitted to the EMBL Data Library, August 1998
Description: Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.
Reference number: Z14675
Accession: T02487
Status: translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-386 <ROU>

A;Cross-references: EMBL:AC004680; NID:g3420043; PIDN:AAC31854.1; PID:g3420053
A;Experimental source: cultivar Columbia
R;lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
W.; Koo, H.; Moffat, K.S.; Cronlin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
Eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, C.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84703
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-386 <STO>
A;Cross-references: GB:AE002093; NID:g3420053; PIDN:AAC31854.1; GSPDB:GM00139
C;Comment: The omega-6 and omega-3 fatty acid desaturases introduce the second and the t
ends of plant membranes.
C;Genetics:
A;Gene: At2g29980; F23F1.10
A;Map position: 2
A;Introns: 103/2; 133/2; 155/3; 186/3; 248/3; 275/3; 321/3
A;Superfamily: omega-3 fatty acid desaturase
C;Keywords: oxidoreductase

Query Match 9.8%; Score 93.5; DB 1; Length 386;
Best Local Similarity 26.5%; Pred. No. 0.13;
Matches 27; Conservative 17; Mismatches 31; Indels 27; Gaps 5;

QY 44 RKAPEVNDLYAVPAIVSIALIYFGSTGIMPLOWIGAGMTAYGLLYPMVHD----- 95
||| : : : : : ||| : : : : :
Db 54 RMSVYVVDIIAV--ALATAAVYVDSWFLNPLYYAAQG-TLFWAIFVLGHDCGHSFSD 110
||| : : : : : ||| : : : : :
QY 96 -----GLVHQRPYPRYPRKGYLKLRYMAHRMHHAVRG 128
||| : : : : : ||| : : : : :
Db 111 IPELLNSVVGHILHS-----FILVPYHGW-----RISHRTHQNHG 145
||| : : : : : ||| : : : : :

RESULT 9
JC7872
stearoyl-CoA 9-desaturase (EC 1.14.19.1), FAD3 - Chlorella vulgaris
N;Alternate names: acyl-CoA desaturase; delta9-desaturase; omega-3 stearoyl-CoA desatura
C;Species: Chlorella vulgaris
C;Date: 09-Dec-2002 #sequence_revision 09-Dec-2002 #text_change 31-Mar-2003
C;Accession: JC7872
R;Suga, K.; Honjo, K.; Furuya, N.; Shimizu, H.; Nishi, K.; Shinohara, F.; Hirabaru, Y.;
Biosci. Biotechnol. Biochem. 66, 1314-1327, 2002
A;Title: Two low-temperature-inducible Chlorella genes for delta12 and omega-3 fatty aci
s cerevisiae, and expression of omega-3 fed in Nicotiana tabacum.
A;Reference number: JC7871; MUID:22152188; PMID:12162554
A;Accession: JC7872
A;Molecule type: mRNA
A;Residues: 1-418 <SUG>
A;Cross-references: DDBJ:AB075527
C;Comment: This enzyme is involved in low temperature adaptation and the development of
C;Genetics:
A;Gene: fad3
C;Keywords: oxidoreductase

Query Match 9.7%; Score 93; DB 2; Length 418;
Best Local Similarity 23.1%; Pred. No. 0.16;
Matches 36; Conservative 19; Mismatches 53; Indels 48; Gaps 6;

QY 57 VPAIVSIALIYFGSTGIMPLOWIGAGMTAYGLLYPMVHDGLVHQRPYPR-- 105
||| : : : : : ||| : : : : :
Db 102 IVAALAGAYTIGNPLVWPLYWFLQG-TMFWALFVVGHD-CGHQSWNNKTLNDFVGNIV 159
||| : : : : : ||| : : : : :
QY 106 ----YIPKGYLKLRYMAHRMHHAVRG-----KEGCVSFGFL 138
||| : : : : : ||| : : : : :
Db 160 HSSIMVPHGW-----RISHRTHHANGHVENDESWHPVIVKSNYEKLQKWLKGLLPPF- 214
||| : : : : : ||| : : : : :
QY 139 YAPPLSKLQATLRPHAAARSAAARDEQGVDTSSSG 174
||| : : : : : ||| : : : : :
Db 215 ---PLFAYPFVLLNRSCKNGSHYDPKSDLFTASEG 247
||| : : : : : ||| : : : : :

RESULT 10
VF2615
conserved hypothetical protein Atu0318 [imported] - Agrobacterium tumefaciens (strain C58)
;Species: Agrobacterium tumefaciens
;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
;Accession: AF2615
;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
;Karp, P.; Romero, P.; Zhang, S.
;Cience 294, 2317-2323, 2001
;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
;ter, E.W
;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
;Reference number: AB2577; MUID:21608550; PMID:11743193
;Accession: AF2615
;Status: preliminary
;Molecule type: DNA
;Residues: 1-304 <KUR>
;Cross-references: GB:AE008698; PIDN:AA141340.1; PID:gi7738653; GSPDB:GN00186
;Experimental source: strain C58 (Dupont)
;Genetics:
;Gene: Atu0318
;Map position: circular chromosome

Query Match 9.5%; Score 91; DB 2; Length 304;
Best Local Similarity 26.8%; Pred. No. 0.18; Mismatches 19; Gaps 9;
Matches 38; Conservative 19; Indels 52; Length 304; Gaps 9;

Y 7 ALIVFTVVGMEVVAALAHKYMVGWGWGHLSHHE-----PRKGAPEVND-LYAVVFA 59

b 125 ALAVTTAEFGLYWAHRIAHTVFF---WRFHAIHSHVRLVWVNTGRFHVADSLFKITLS 181

Y 60 IVSIALIYFGSTGIWPLQ---WTGAGTAYGLLYFMVHDGLVHQWPPRYI---PRKGYL 113

b 182 --QIPLYFMGA-----PLQVFMIGAVTAFIGIL---THCNVDMKTGLLDYIFSTPR---- 228

Y 114 KRLYMAHRMHAHVGRGCVSFG 136

b 229 -----LHRWHSKQLPBGNTNYG 246

RESULT 11
97397

Ypohetical protein AGR_C_556 [imported] - Agrobacterium tumefaciens (strain C58), Cere
;Species: Agrobacterium tumefaciens
;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
;Accession: E97397
;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
;A.; Liu, F.; Woliam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
;Cience 294, 2323-2328, 2001
;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
;Reference number: A97359; MUID:21608551; PMID:11743194
;Accession: E97397
;Status: preliminary
;Molecule type: DNA
;Residues: 1-313 <KUR>
;Cross-references: GB:AE007869; PIDN:AAK86134.1; PID:gi5155221; GSPDB:GN00169
;Genetics:
;Gene: AGR_C_556
;Map position: circular chromosome

Query Match 9.5%; Score 91; DB 2; Length 313;
Best Local Similarity 26.8%; Pred. No. 0.18; Mismatches 19; Gaps 9;
Matches 38; Conservative 19; Indels 52; Length 313; Gaps 9;

Y 7 ALIVFTVVGMEVVAALAHKYMVGWGWGHLSHHE-----PRKGAPEVND-LYAVVFA 59

b 134 ALAVTTAEFGLYWAHRIAHTVFF---WRFHAIHSHVRLVWVNTGRFHVADSLFKITLS 190

Y 60 IVSIALIYFGSTGIWPLQ---WTGAGTAYGLLYFMVHDGLVHQWPPRYI---PRKGYL 113

b 191 --QIPLYFMGA-----PLQVFMIGAVTAFIGIL---THCNVDMKTGLLDYIFSTPR---- 237

Qy 114 KRLYMAHRMHAHVGRGCVSFG 136

Db 238 -----LHRWHSKQLPBGNTNYG 255

RESULT 12

G83556

hypothetical protein PA0702 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83556

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.K.; Kas, A.; Larbig, K.; Lim

;- Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83556

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <STO>

A:Cross-references: GB:AE004506; GB:AE004091; NID:g9946584; PIDN:AAG04091.1; GSPDB:GN00

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA0702

Query Match 9.2%; Score 88; DB 2; Length 301;
Best Local Similarity 24.8%; Pred. No. 0.35; Mismatches 24; Indels 20; Gaps 6;
Matches 36; Conservative 24; Mismatches 65; Indels 20; Gaps 6;

Qy 2 LWNALIVFTVVGMEVVAALAHKYMVGWGWGHLSHHEPRKGAPEVNDLY---AVVF 58

Db 106 LMLQLLAIVLADLIGITVHYASHRSALL---WRLHAVHH-----SVQRLYGFNGLMK 155

Qy 59 AIVSIALIYFGSTGIWPLQ---WTGAGTAYGLLYFMVHDGLVHQWPPRYIPRKGYLKLYM 118

Db 156 HPLHLGLEALG--GTLPLLLGVPQTVAALLAFAIGIQLLQF--SNVDMRTGGJLHVFA 211

Qy 119 ---AHRMHAHVGRGCVSFGFLYA 140

Db 212 WAPLRLHRLHRIYGRAGDVNFALFFS 236

RESULT 13

S72765

Phosphoenolpyruvate carboxylase (EC 4.1.1.31) ppc - Mycobacterium leprae

N:Alternate names: B1496 C3 207

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72765

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B1496.

A:Reference number: S72695

A:Accession: S72765

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 <SMI>

A:Cross-references: EMBL:U00013; NID:g466868; PIDN:AAA17132.1; PID:g466887

C:Genetics:

A:Gene: ppc

C:Superfamily: phosphoenolpyruvate carboxylase

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 9.1%; Score 87; DB 2; Length 907;
Best Local Similarity 23.5%; Pred. No. 1.4; Mismatches 21; Indels 53; Gaps 7;
Matches 38; Conservative 21; Mismatches 53; Indels 50; Gaps 7;

Qy 44 RKGAPEVNDLYAVVFAIV---SIALIYFGSTGIWPLQ---WTGAGTAYGLLYFMVHDGLVH 99

Db 725 RKPTTSIALDLRAIPWLAWSQSRVMLPGWYGGSAFQWVAAGPSESQRVEMLHD---LY 782

Qy 100 QRWPF-----RYIP-----RKGYLKLYMAHR-----MHAVR 127

|||||
783 QRWPFRRSVLSNMAQVLAKSDGLAARYAELVWDEALRRRVEFKIADERRRTIAIHKLIT 842
128 GKGCVSFG-----FLYAPPLSKLQATLRERHAARG 159
843 GHDDLLADNPALARSVFNRPFPYLEPLNHLQVELLRRY--RSG 882

SULT 14
6981
obable phosphoenolpyruvate carboxylase [imported] - Mycobacterium leprae
Species: Mycobacterium leprae
Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
Accession: B86981
Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
m, M.A.; Rutherford, K.M.
ture 409, 1007-1011, 2001
Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
Title: Massive gene decay in the leprosy bacillus.
Reference number: A86909; MUID:21128732; PMID:11234002
Accession: B86981
Status: preliminary
Molecule type: DNA
Residues: 1-934 <STO>
Cross-references: GB:AL450380; NID:gl3092776; PIDN:CA30086.1; GSPDB:GN00147
Genetics:
Gene: ppc
Superfamily: phosphoenolpyruvate carboxylase

Query Match 9.1%; Score 87; DB 2; Length 934;
Best Local Similarity 23.5%; Pred.No. 1.5;
Matches 38; Conservative 21; Mismatches 53; Indels 50; Gaps 7;
44 RKGAPEVNDLYAVFAIV-----SIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHEDGLVH 99
752 RKPTTSTADLRAPWVLANSQSRWMLPGWYGTGSAFQWVAQPESESQVEMLED--LY 809
100 QRWPF-----RYIP-----RKGYLKLRYMAHR-----MHAVR 127
810 QRWPFRRSVLSNMAQVLAKSDGLAARYAELVWDEALRRRVEFKIADERRRTIAIHKLIT 869

128 GKGCVSFG-----FLYAPPLSKLQATLRERHAARG 159
870 GHDDLLADNPALARSVFNRPFPYLEPLNHLQVELLRRY--RSG 909

SULT 15
3586
pothetical membrane spanning protein BMEII0616 [imported] - Brucella melitensis (strain
Species: Brucella melitensis
Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
Accession: AG3586
DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
Mazur, M.; Gotsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
oc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
Reference number: AD3252; PMID:11756688
Accession: AG3586
Status: preliminary
Molecule type: DNA
Residues: 1-162 <KUR>
Cross-references: GB:AE008918; PIDN:AAL53858.1; PID:gi7994795; GSPDB:GN00191
Experimental source: strain 16M
Genetics:
Gene: BMEII0616
Map position: II

Query Match 8.8%; Score 84.5; DB 2; Length 162;
Best Local Similarity 27.0%; Pred.No. 0.41;
Matches 41; Conservative 23; Mismatches 41; Indels 47; Gaps 13;
4 INNALIVFV--TVVGMEVVAALAHKYNHMG-----WG-----WGMHLSR-----HE 42

|||||
5 IWRGILIGIGATVV-NDIWAQLL---ALLPGSRPKNGLVGRWFHLEHGOIFHDDISQSE 61
43 PRK-----GAFVNDLYAVFAIVSIALYFGSTGIWPLQWIGAGMTAYGLL----- 89
62 PCRHEVALGWTGYAVGILYGVIFALYGGA-AWFAFIFLP-AWI-----FGILTIAAG 113
90 YFMVHVDGLVHQRWPFYIPRKGYLKLRL-YMAH 120
114 WFLQPGI-GIGWAASKLPNAGNVRILNLIAH 144

Search completed: February 29, 2004, 14:52:57
Job time : 7.87104 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

[protein - protein search, using sw model

in on: February 29, 2004, 14:51:24 ; Search time 11.619 seconds
(without alignments)
3180.293 Million cell updates/sec

file: US-09-941-947A-36

Effect score: 956

Sequence: 1 MLMTWNLALIVFTVVGMEVV.....ARSCAARDGGQVDTSSGK 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Aligned: 809742 seqs, 211153259 residues

total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	956	100.0	175	10	US-09-941-947A-36
2	956	100.0	175	14	US-10-115-571A-12
3	892	93.3	175	9	US-09-323-998D-7
4	854	88.4	175	9	US-09-323-998D-6
5	499	52.2	162	9	US-09-323-998D-5
6	495	51.8	162	10	US-09-920-923-34
7	488	51.0	162	9	US-09-323-998D-8
8	487	50.9	169	9	US-09-547-267-9
9	487	50.9	169	14	US-09-920-923-6
10	487	50.9	169	10	US-10-166-225A-183
11	247	25.8	294	9	US-09-323-998D-4
12	245	25.6	309	15	US-10-259-194A-398
13	101.5	10.6	383	14	US-10-115-571A-3
14	101.5	10.6	383	14	US-10-115-571A-10
15	100.5	10.5	380	14	US-10-115-571A-1

Sequence 34, Appl
Sequence 2, Appl
Sequence 9, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 6, Appl
Sequence 16, Appl
Sequence 3466, Appl
Sequence 11607, A
Sequence 15, Appl
Sequence 458, Appl
Sequence 19, Appl
Sequence 28, Appl
Sequence 14, Appl
Sequence 246, Appl
Sequence 3871, Appl
Sequence 18, Appl
Sequence 20, Appl
Sequence 21, Appl
Sequence 59, Appl
Sequence 467, Appl
Sequence 37, Appl
Sequence 17, Appl
Sequence 3762, Appl
Sequence 160, Appl
Sequence 23, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 14555, A
Sequence 30, Appl

16 99.5 10.4 377 9 US-09-837-751-34
17 99.5 10.4 377 14 US-10-115-571A-2
18 99.5 10.4 377 14 US-10-115-571A-9
19 93.5 8.8 386 14 US-10-115-571A-4
20 93.5 9.8 386 14 US-10-115-571A-11
21 92 9.6 379 14 US-10-115-571A-68
22 91.5 9.5 358 14 US-10-115-571A-16
23 90.5 9.5 549 15 US-10-369-493-3466
24 87 9.1 335 14 US-10-156-761-11607
25 86.5 9.0 351 14 US-10-115-571A-15
26 86.5 9.0 435 15 US-10-310-154-468
27 86 9.0 352 14 US-10-115-571A-19
28 84.5 8.8 360 14 US-10-115-571A-26
29 84 8.8 375 14 US-10-115-571A-14
30 83 8.7 350 15 US-10-295-027-246
31 83 8.7 413 15 US-10-369-493-3871
32 82.5 8.6 362 14 US-10-115-571A-18
33 82.5 8.6 363 14 US-10-115-571A-20
34 82.5 8.6 369 14 US-10-115-571A-21
35 82.5 8.6 382 14 US-10-115-571A-69
36 82.5 8.6 446 15 US-10-310-154-467
37 81.5 8.5 321 14 US-10-115-571A-37
38 81.5 8.5 377 14 US-10-115-571A-17
39 81.5 8.5 429 15 US-10-369-493-3762
40 81 8.5 347 15 US-10-259-194A-160
41 81 8.5 368 14 US-10-115-571A-23
42 80 8.4 371 9 US-09-885-189-10
43 80 8.4 372 9 US-09-885-189-10
44 79 8.3 269 14 US-10-156-761-14555
45 78.5 8.2 362 14 US-10-115-571A-30

ALIGNMENTS

RESULT 1

US-09-941-947A-36
; Sequence 36, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Bzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: Dicosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOLD PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941.947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 175
; TYPE: PPT
; ORGANISM: Pantoea stewartii
US-09-941-947A-36

Query Match 100.0%; Score 956; DB 10; Length 175;
Best Local Similarity 100.0%; Pred. No. 9e-101;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLMTWNLALIVFTVVGMEVVAALAHKYIMEGWGWHLSHHEPRKGAFVNDLYAVVFAI 60

DB 1 MLMTWNLALIVFTVVGMEVVAALAHKYIMEGWGWHLSHHEPRKGAFVNDLYAVVFAI 60

QY 61 VSIALIVFGSTGIPLOWIGAGMTAYGLLYFMVDGLVHQWPFYIPRKGLKRLTMAH 120

b 61 VSTALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120
y 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 175
b 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 175

RESULT 2

S-10-218-118-12
Sequence 12, Application US/10218118
Publication No. US2003014831A1
GENERAL INFORMATION:
APPLICANT: Bryostowicz, Patricia
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
FILE REFERENCE: CL1876 US NA
CURRENT APPLICATION NUMBER: US/10/218,118
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: 60/312,646
PRIOR FILING DATE: 2001-08-15
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 12
LENGTH: 175
TYPE: PRT
ORGANISM: Pastoea stewartii
S-10-218-118-12

Query Match 100.0%; Score 956; DB 14; Length 175;
Best Local Similarity 100.0%; Pred. No. 9e-101;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60
b 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60
y 61 VSTALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120
b 61 VSTALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120
y 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 175
b 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 175

RESULT 3

S-09-323-998D-7
Sequence 7, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/624,125
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 175
TYPE: PRT

; ORGANISM: Erwinia uredovora
US-09-323-998D-7
Query Match 93.3%; Score 892; DB 9; Length 175;
Best Local Similarity 91.4%; Pred. No. 1.8e-93;
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
y 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60
b 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60
y 61 VSTALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120
b 61 VSTALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120
y 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 175
b 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 175

RESULT 4

US-09-323-998D-6
Sequence 6, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/624,125
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 175
TYPE: PRT
ORGANISM: Erwinia herbicola
US-09-323-998D-6

Query Match 68.4%; Score 654; DB 9; Length 175;
Best Local Similarity 68.2%; Pred. No. 2.3e-66;
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;
y 6 NALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 65
b 3 NSLIVLSVTAMEGIAAFTHRYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 62
y 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 125
b 63 IAYGTAGVWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 122
y 126 VRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 173
b 123 VRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEQDGVDTSSGK 172

RESULT 5

US-09-323-998D-5
Sequence 5, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF

APPLICATION NUMBER: US/09/547,267

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 169 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

S-09-547-267-9

Query Match 50.9%; Score 487; DB 9; Length 169;

Best Local Similarity 51.5%; Pred. No. 2.3e-47;

Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;

Y 5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGNGWHLSSH-EPRKGAFVNDLYAVVFAIV 61

b 4 WAAILTVILTVAAAMELTAVSVHVMHGPGLGNGWHSKSHDEHDHDALEKNDLYGVIFAVI 63

Y 62 SIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVDGLVHQRWPFYIPRKYGLKRLYMAHR 121

b 64 SIVLFAIGMGSDLAWLAVGVTCYGLIYFLHDGLVHGRWPFYVPRKYGLRVRVQHR 123

Y 122 MHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGVDTSS 172

b 124 MHAVHGRENCVSFGFIWAPSVDSLAKELK-----RSGALLKDREGADNT 169

RESULT 9

S-09-920-923-6

Sequence 6, Application US/09920923

Publication No. US2003002273A1

GENERAL INFORMATION:

APPLICANT: Paeamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/09/920,923

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832

PRIOR FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 6

LENGTH: 169

TYPE: PRT

ORGANISM: *Plavobacterium* sp. R1534

S-09-920-923-6

Query Match 50.9%; Score 487; DB 10; Length 169;

Best Local Similarity 51.5%; Pred. No. 2.3e-47;

Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;

Y 5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGNGWHLSSH-EPRKGAFVNDLYAVVFAIV 61

b 4 WAAILTVILTVAAAMELTAVSVHVMHGPGLGNGWHSKSHDEHDHDALEKNDLYGVIFAVI 63

Y 62 SIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVDGLVHQRWPFYIPRKYGLKRLYMAHR 121

b 64 SIVLFAIGMGSDLAWLAVGVTCYGLIYFLHDGLVHGRWPFYVPRKYGLRVRVQHR 123

Y 122 MHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGVDTSS 172

Db 124 MHAVHGRENCVSFGFIWAPSVDSLAKELK-----RSGALLKDREGADNT 169

RESULT 10

US-10-166-225A-183

Sequence 183, Application US/10166225A

Publication No. US20030148416A1

GENERAL INFORMATION:

APPLICANT: BERRY, Alan

APPLICANT: BRETZEL, Werner

APPLICANT: HUMBELIN, Markus

APPLICANT: IOPEZ-ULIBARRI, Rual

APPLICANT: MAYER, Anne F.

APPLICANT: YELISEEV, Alexei A.

TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION

FILE REFERENCE: C38435/121966

CURRENT APPLICATION NUMBER: US/10/166,225A

CURRENT FILING DATE: 2002-06-05

NUMBER OF SEQ ID NOS: 197

SOFTWARE: Patent in version 3.1

SEQ ID NO 183

LENGTH: 169

TYPE: PRT

ORGANISM: *Paracoccus* sp. R1534

US-10-166-225A-183

Query Match 50.9%; Score 487; DB 14; Length 169;

Best Local Similarity 51.5%; Pred. No. 2.3e-47;

Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;

QY 5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGNGWHLSSH-EPRKGAFVNDLYAVVFAIV 61

Db 4 WAAILTVILTVAAAMELTAVSVHVMHGPGLGNGWHSKSHDEHDHDALEKNDLYGVIFAVI 63

QY 62 SIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVDGLVHQRWPFYIPRKYGLKRLYMAHR 121

Db 64 SIVLFAIGMGSDLAWLAVGVTCYGLIYFLHDGLVHGRWPFYVPRKYGLRVRVQHR 123

QY 122 MHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGVDTSS 172

Db 124 MHAVHGRENCVSFGFIWAPSVDSLAKELK-----RSGALLKDREGADNT 169

RESULT 11

US-09-323-998D-4

Sequence 4, Application US/09323998D

Patent No. US20020102631A1

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIREN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND

METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998D

CURRENT FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 08/937,155

PRIOR FILING DATE: 1997-09-25

PRIOR APPLICATION NUMBER: 05/624,125

PRIOR FILING DATE: 1996-03-29

NUMBER OF SEQ ID NOS: 61

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 4

LENGTH: 294

TYPE: PRT

ORGANISM: *Arabidopsis thaliana*

US-09-323-998D-4

Query Match 25.8%; Score 247; DB 9; Length 294;

Best Local Similarity 41.2%; Pred. No. 1e-19;
Matches 54; Conservative 24; Mismatches 47; Indels 6; Gaps 3;
15 VGEVVAALAHKYMIGWGWHLSHREPKGAFVNDLVAVVFAIVSIALIYFG--STG 72
131 VGEVFWARWARALWHLASLNNHSHKHPREGFELNDVFAIVNAGPAIGLLSYGPFNG 190
73 IWLQWAGM--TAYGLLYFVMDGLVHQRWPFYIPRKGYLKRLYMAHRMHAHVGRKE 130
191 LVPGCEGAGLIGTVFGIAYMFVHDGLVHRKFRFVGPIDVPLRVKVAARFQLHHT--DKF 248
131 GCVSFGFLYAP 141
249 NGVPYGLFLGP 259

SULT 12
-10-259-194A-398
Sequence 398, Application US/10259194A
Publication No. US20040010815A1
GENERAL INFORMATION:

APPLICANT: Lange, Markus B.
APPLICANT: Ghassseman, Majid
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazebrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Moughamer, Todd
APPLICANT: Provart, Nicholas
APPLICANT: Ricke, Darrell
APPLICANT: Zhu, Tong

TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

FILE REFERENCE: 70029-NP
CURRENT APPLICATION NUMBER: US/10/259, 194A
CURRENT FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,743
PRIOR FILING DATE: 2002-04-04
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 662
SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
SEQ ID NO 398
LENGTH: 309
TYPE: PRT
ORGANISM: Oryza sativa
-10-259-194A-398

Query Match 25.6%; Score 245; DB 15; Length 309;
Best Local Similarity 41.2%; Pred. No. 1.8e-19;
Matches 54; Conservative 23; Mismatches 48; Indels 6; Gaps 4;
15 VGEVVAALAHKYMIGWGWHLSHREPKGAFVNDLVAVVFAIVSIALIYFG--STG 72
158 VGKFWAQAHRSLWHLASLNNHSHKHPREGFELNDVFAIVNAGPAIGLLSYGPFNG 217
73 IWLQWAGM--TAYGLLYFVMDGLVHQRWPFYIPRKGYLKRLYMAHRMHAHVGRKE 130
218 IVFGLCEGAGLIGTILGMAVYFVMDGLVHRFPVGPIDVPLRVKVAARFQLHHTOK-FE 276
131 GCVSFGFLYAP 141
277 G-VPYGLFLGP 286

SULT 13
-10-115-571A-3
Sequence 3, Application US/10115571A
Publication No. US20030150020A1
GENERAL INFORMATION:

APPLICANT: HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF
APPLICANT: AGRICULTURE AND AGRI-FOOD CANADA
APPLICANT: SOMERS, Daryl
APPLICANT: RAKOW, Gerhard
TITLE OF INVENTION: PLANT FATTY ACID DESATURASES AND ALLELES THEREFOR
FILE REFERENCE: 81601-28
CURRENT APPLICATION NUMBER: US/10/115,571A
CURRENT FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: CA 2,284,246
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 383
TYPE: PRT
ORGANISM: Brassica napus
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: GenBank/L01418
DATABASE ENTRY DATE: 1993-04-27
US-10-115-571A-3

Query Match 10.6%; Score 101.5; DB 14; Length 383;
Best Local Similarity 25.5%; Pred. No. 0.0052;
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;

QY 44 RKGAFEVNDLYAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHD----- 95
Db 51 RSMYSYTRDIFAV--AALAAVYFDSWFLWPLVYVAQG-TLFWAIFVLGHDGCHGSPSD 107
QY 96 -----GLVHQRWPFYIPRKGYLKRLYMAHRMHAHVGR 128
Db 108 IPLNSVVGHLHS---FILVPYHGW----RISHRTHQKHG 142

RESULT 14

US-10-115-571A-10
Sequence 10, Application US/10115571A
Publication No. US20030150020A1
GENERAL INFORMATION:
APPLICANT: HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF
APPLICANT: AGRICULTURE AND AGRI-FOOD CANADA
APPLICANT: SOMERS, Daryl
APPLICANT: RAKOW, Gerhard
TITLE OF INVENTION: PLANT FATTY ACID DESATURASES AND ALLELES THEREFOR
FILE REFERENCE: 81601-28
CURRENT APPLICATION NUMBER: US/10/115,571A
CURRENT FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: CA 2,284,246
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 383
TYPE: PRT
ORGANISM: Brassica napus
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: GenBank/P48624
DATABASE ENTRY DATE: 1996-02-01
RELEVANT RESIDUES: (1) ..(383)
US-10-115-571A-10

Query Match 10.6%; Score 101.5; DB 14; Length 383;
Best Local Similarity 25.5%; Pred. No. 0.0052;
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;

QY 44 RKGAFEVNDLYAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHD----- 95
Db 51 RSMYSYTRDIFAV--AALAAVYFDSWFLWPLVYVAQG-TLFWAIFVLGHDGCHGSPSD 107
QY 96 -----GLVHQRWPFYIPRKGYLKRLYMAHRMHAHVGR 128
Db 108 IPLNSVVGHLHS---FILVPYHGW----RISHRTHQKHG 142

RESULT 15
S-10-115-571A-1
Sequence 1, Application US/10115571A
Publication No. US20030150020A1
GENERAL INFORMATION:
APPLICANT: HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF
AGRICULTURE AND AGRI-FOOD CANADA
APPLICANT: SOMERS, Daryl
APPLICANT: RAKOW, Gerhard
TITLE OF INVENTION: PLANT FATTY ACID DESATURASES AND ALLELES THEREFOR
FILE REFERENCE: 81601-28
CURRENT APPLICATION NUMBER: US/10/115,571A
CURRENT FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: CA 2,284,246
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 380
TYPE: PRT
ORGANISM: Brassica napus
FEATURE:
NAME/KEY: SITE
LOCATION: (95)..(99)
OTHER INFORMATION: Histadine box sequence
FEATURE:
NAME/KEY: SITE
LOCATION: (213)..(213)
OTHER INFORMATION: Position of amino acid substitution in accordance with various as
OTHER INFORMATION: pects of the invention
FEATURE:
NAME/KEY: SITE
LOCATION: (275)..(275)
OTHER INFORMATION: Position of amino acid substitution in accordance with various as
OTHER INFORMATION: pects of the invention
FEATURE:
NAME/KEY: SITE
LOCATION: (347)..(347)
OTHER INFORMATION: Position of amino acid substitution in accordance with various as
OTHER INFORMATION: pects of the invention
S-10-115-571A-1
Query Match 10.5%; Score 100.5; DB 14; Length 380;
Best Local Similarity 26.3%; Pred. No. 0.0067;
Matches 26; Conservative 17; Mismatches 35; Indels 21; Gaps 4;
ly 44 RKGAFEVNDLYAVVFAIVSIALIYFGSTGIWPLQWIGAGTAYGLLYFMVHD----- 95
lb 48 RMSYVARDIESVV--ALAVAVYFDSWFFWPLYMAAQG-TLFWALFVLGHDCGHGFSFD 104
ly 96 -----GLVHQRPFRYIPKGYLKYLYMAHRMHAVRG 128
lb 105 IPLLNTAVGHILHSFILVPYHGW----RMSRTHQNHG 139
Search completed: February 29, 2004, 15:28:06
Ob time : 12.619 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 5.31793 Seconds
(without alignments)
1698.885 Million cell updates/sec

le: US-09-941-947a-36

fect score: 956

quence: 1 MLMTWNLIVFTVVGMEV.....ARSGAARDEQDGYDTSSGK 175

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

imum DB seq length: 0

ximum DB seq length: 2600000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA.*

- 1: /cgm2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgm2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgm2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgm2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgm2_6/ptodata/2/iaa/PTUS_COMB.pep.*
- 6: /cgm2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	892	93.3	175	1	US-07-783-705A-6
2	892	93.3	175	1	US-08-624-125-7
3	892	93.3	175	4	US-08-937-155-7
4	892	93.3	175	4	US-09-323-998E-7
5	654	68.4	175	1	US-08-624-125-6
6	654	68.4	175	4	US-08-937-155-6
7	654	68.4	175	4	US-09-323-998E-6
8	653	68.3	176	1	US-08-096-623A-18
9	499	52.2	162	1	US-08-624-125-5
10	499	52.2	162	2	US-08-663-310-11
11	499	52.2	162	2	US-09-006-491-11
12	499	52.2	162	3	US-09-335-919-11
13	499	52.2	162	4	US-08-937-155-5
14	499	52.2	162	4	US-09-323-998E-5
15	495	51.8	162	2	US-08-663-310-4
16	495	51.8	162	2	US-09-006-491-4
17	495	51.8	162	3	US-09-335-919-4
18	495	51.8	162	3	US-08-980-832-34
19	495	51.8	162	4	US-09-920-923B-34
20	488	51.0	162	1	US-08-624-125-8
21	488	51.0	162	4	US-08-937-155-8
22	488	51.0	162	4	US-09-323-998E-8
23	487	50.9	169	3	US-08-660-645A-9
24	487	50.9	169	3	US-09-298-718-9
25	487	50.9	169	3	US-09-546-969-9
26	487	50.9	169	3	US-08-980-832-6
27	487	50.9	169	4	US-09-547-267-9

Sequence 6, Appli
Sequence 14, Appli
Sequence 2, Appli
Sequence 15, Appli
Sequence 16, Appli
Sequence 4, Appli
Sequence 10, Appli
Sequence 13, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 12, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 8, Appli
Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-07-783-705A-6

; Sequence 6, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939iniko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, Shigeo
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: N/A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/783,705A
FILING DATE: 19911023
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-103078
FILING DATE: 21-APR-1989
APPLICATION NUMBER: JP 2-53225
FILING DATE: 05-MAR-1990
APPLICATION NUMBER: US 07/519,011
FILING DATE: 19-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schwadron, Janet I.
REGISTRATION NUMBER: 33,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-708-1935
TELEFAX: 212-246-5959
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-705A-6

Query Match 93.3%; Score 892; DB 1; Length 175;

Best Local Similarity 91.4%; Pred. No. 1.1e-101;
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Y 1 MLEIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60
b 1 MLEIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAA 60

Y 61 VSIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPFYIPRKGYLKLYMAH 120
b 61 LSILLIYLGSTGMPWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPFYIPRKGYLKLYMAH 120

Y 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175
b 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

RESULT 2
S-08-624-125-7
Sequence 7, Application US/08624125
Patent No. 5744341
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,125
FILING DATE: 29-MAR-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-624-125-7
Query Match 93.3%; Score 892; DB 1; Length 175;
Best Local Similarity 91.4%; Pred. No. 1.1e-101;
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Y 1 MLEIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60
b 1 MLEIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAA 60

Y 61 VSIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPFYIPRKGYLKLYMAH 120
b 61 LSILLIYLGSTGMPWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPFYIPRKGYLKLYMAH 120

Y 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175
b 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

Db 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

RESULT 3
US-08-937-155-7
Sequence 7, Application US/08937155
Patent No. 6524811
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT.
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,155
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/624,125
FILING DATE: 29-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-937-155-7
Query Match 93.3%; Score 892; DB 4; Length 175;
Best Local Similarity 91.4%; Pred. No. 1.1e-101;
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLEIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60
Db 1 MLEIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAA 60

QY 61 VSIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPFYIPRKGYLKLYMAH 120
Db 61 LSILLIYLGSTGMPWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPFYIPRKGYLKLYMAH 120

QY 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175
Db 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

RESULT 4
US-09-323-998E-7
Sequence 7, Application US/09323998E
Patent No. 6642021
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998E
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 175
TYPE: PRT
ORGANISM: Erwinia uredovora
-09-323-998B-7

Query Match 93.3%; Score 892; DB 4; Length 175;
Best Local Similarity 91.4%; Pred. No. 1.1e-101; Indels 0; Gaps 0;
Matches 160; Conservative 6; Mismatches 9;

1 MLMTWNAIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAI 60
|||||
1 MLMTWNAIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPA 60
|||||
61 VSIALIYFGSTGIPWLOIGAGMTAYGLLYFMVHDGLVHQRPPIPRKGYLKRLYMAH 120
|||||
61 LSILLIYLGSTGMPWLOIGAGMTAYGLLYFMVHDGLVHQRPPIPRKGYLKRLYMAH 120
|||||
121 RMHVAVRGKGVSGFLYAPPLSKQATLRERHAARSAGARDEQGVDTSSGK 175
|||||
121 RMHVAVRGKGVSGFLYAPPLSKQATLRERHAGARDAQGGEDBPASGK 175
|||||

SULT 5
-08-624-125-6
Sequence 6, Application US/08624125
Patent No. 5744341
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,125
FILING DATE: 29-MAR-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/624,125
FILING DATE: 29-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-624-125-6

Query Match 68.4%; Score 654; DB 1; Length 175;
Best Local Similarity 68.2%; Pred. No. 2e-72;
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

6 NALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAI 65
|||||
3 NSLIVILSVIAMEGIAAFTTHRYIMHGNGWGHLSHHEPRKGVFELNDLPVAVFAGVAIAL 62
|||||
66 IYFGSTGIPWLOIGAGMTAYGLLYFMVHDGLVHQRPPIPRKGYLKRLYMAH 125
|||||
63 IAVGTAGVPLQWIGCGMTYGLLYFLVHDGLVHQRPPIPRKGYLKRLYMAH 122
|||||
126 VRGKGVSGFLYAPPLSKQATLRERHA--ARSGAARDEQGVDTSSS 173
|||||
123 VRGKGVSGFLYAPPLSKQATLRERHGRPPKRDAAKDRPDAAAPSSS 172
|||||

RESULT 6
US-08-937-155-6
Sequence 6, Application US/08937155
Patent No. 6524811
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,155
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/624,125
FILING DATE: 29-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-937-155-6

Query Match 68.4%; Score 654; DB 4; Length 175;
Best Local Similarity 68.2%; Pred. No. 2e-72;
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

6 NALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAI 65
|||||
3 NSLIVILSVIAMEGIAAFTTHRYIMHGNGWGHLSHHEPRKGVFELNDLPVAVFAGVAIAL 62
|||||

D 3 NSLIVLTVIAMEGIAAFTHYIMEGWGRWHESHHTPRKGVFLNDLFAVVFAGVIAL 62
Y 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAHRMHA 125
C 63 IAVGTAGVWPLQWIGCGMTVYGLLYFLVHDGLVHORWPFYIPRKGYLKRLYVAHRLHA 122
Y 126 VRGKGCVSFGFLYAPPLSKLQATLRERHA--ARSGAARDEODGVDTSSS 173
C 123 VRGKGCVSFGFIYARKPADLQAILRERHGRPPKRDAAKRDPAASPSS 172

RESULT 7

S-09-323-998E-6

Sequence 6, Application US/09323998E

Patent No. 6642021

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIREN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND

METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998E

PRIOR FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1998-06-02

NUMBER OF SEQ ID NOS: 62

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 6

LENGTH: 175

TYPE: PRT

ORGANISM: Erwinia herbicola

S-09-323-998E-6

Query Match 68.4%; Score 654; DB 4; Length 175;
Best Local Similarity 68.2%; Pred. NO. 2e-72;
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

Y 6 NALIVFTVVGMEVVAALAHKYIMEGWGWHLSHHEPRKGAFVNDLYAVVFATVIAL 65
b 3 NSLIVLTVIAMEGIAAFTHYIMEGWGRWHESHHTPRKGVFLNDLFAVVFAGVIAL 62
Y 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAHRMHA 125
b 63 IAVGTAGVWPLQWIGCGMTVYGLLYFLVHDGLVHORWPFYIPRKGYLKRLYVAHRLHA 122
Y 126 VRGKGCVSFGFLYAPPLSKLQATLRERHA--ARSGAARDEODGVDTSSS 173
b 123 VRGKGCVSFGFIYARKPADLQAILRERHGRPPKRDAAKRDPAASPSS 172

RESULT 8

S-08-096-623A-18

Sequence 18, Application US/08096623A

Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.

APPLICANT: Brinkhaus, Friedhelm L.

APPLICANT: Mukharji, Indrani

APPLICANT: Proffitt, John H.

APPLICANT: Yarger, James G.

APPLICANT: Yen, Huei-Che B.

TITLE OF INVENTION: Biosynthesis of Zeaxanthin and

Glycosylated Zeaxanthin in Genetically Engineered Hosts

NUMBER OF SEQUENCES: 104

CORRESPONDENCE ADDRESS:

ADDRESSEE: Welch & Katz, Ltd

STREET: 120 S. Riverside Plaza, 22nd Floor

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Ganson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 176 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-623A-18

Query Match 68.3%; Score 653; DB 1; Length 176;
Best Local Similarity 68.2%; Pred. No. 2.7e-72;
Matches 116; Conservative 20; Mismatches 32; Indels 2; Gaps 1;
Qy 6 NALIVFTVVGMEVVAALAHKYIMEGWGWHLSHHEPRKGAFVNDLYAVVFATVIAL 65
Db 4 NSLIVLTVIAMEGIAAFTHYIMEGWGRWHEPHHTPRKGVFLNDLFAVVFAGVIAL 63
Qy 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAHRMHA 125
Db 64 IAVGTAGVWPLQWIGCGMTVYGLLYFLVHDGLVHORWPFYIPRKGYLKRLYVAHRLHA 123
Qy 126 VRGKGCVSFGFLYAPPLSKLQATLRERHA--ARSGAARDEODGVDTSSS 173
Db 124 VRGKGCVSFGFIYARKPADLQAILRERHGRPPKRDAAKRDPAASPSS 173

RESULT 9

US-08-624-125-5

Sequence 5, Application US/08624125

Patent No. 5744341

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIREN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND

METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

RESULT 13
 US-08-937-155-5
 Sequence 5, Application US/08937155
 Patent No. 6524811
 GENERAL INFORMATION:
 APPLICANT: CUNNINGHAM JR., FRANCIS X.
 APPLICANT: SUN ZAIREN
 TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
 TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/937,155
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/624,125
 FILING DATE: 29-MAR-1996
 ATTORNEY/AGENT INFORMATION:

C the carotenoid biosynthetic pathway and which metabolise single carbon
C substrates. The carotenoids have potent anti-oxidant properties useful in
C diet, and aquaculture elements. The carotenoids are also useful as
C intermediates in the synthesis of steroids flavours and fragrances and
C compounds for potential electro-optic applications. The present sequence
C is Pantoea stewartii beta-carotene hydrolase (CrtZ) enzyme used in the
C invention

X Sequence 175 AA;

Query Match 100.0%; Score 956; DB 5; Length 175;
Best Local Similarity 100.0%; Pred. No. 9.7e-108;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60
b 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60
Y 61 VSTALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKLYMAH 120
b 61 VSTALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKLYMAH 120
Y 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEQDGVDTSSGK 175
b 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEQDGVDTSSGK 175

RESULT 2

AO16023
D AAO16023 standard; protein; 175 AA.

X C AAO16023;

X T 20-FEB-2003 (first entry)

X E Pantoea stewartii Beta-carotene hydroxylase.

X W Carotenoid; crt.

X S Pantoea stewartii.

X N W0200279395-A2.

X D 10-OCT-2002.

X F 25-JAN-2002; 2002WO-US0002124.

X R 26-JAN-2001; 2001US-0264329P.

X R 04-MAY-2001; 2001US-0288984P.

X A (CRGI) CARGILL INC.

X X De Souza ML, Kollmann SR, May CA, Schroeder WA;

X I WPI; 2003-075455/07.

X R N-PSDB; RET14195.

X Novel isolated nucleic acid useful e.g. to engineer host cells with the
X ability to produce particular carotenoids and polypeptides useful in cell
X -free systems to make particular carotenoids.

X Claim 38; Page 65-66; 74pp; English.

X The invention comprises the amino acid and coding sequence of a number of
X carotenoid (crt)-related proteins. The crt-related DNA and protein
X sequences of the invention are useful for engineering cells which are
X able to produce carotenoids. The present amino acid sequence represents a
X crt-related protein of the invention

X Sequence 175 AA;

Query Match 100.0%; Score 956; DB 6; Length 175;
Best Local Similarity 100.0%; Pred. No. 9.7e-108;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60
Db 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60
Qy 61 VSTALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKLYMAH 120
Db 61 VSTALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKLYMAH 120
Qy 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEQDGVDTSSGK 175
Db 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEQDGVDTSSGK 175

RESULT 3

ABP96690
ID ABP96690 standard; protein; 175 AA.

XX ABP96690;

XX 03-JUN-2003 (first entry)

XX Pantoea stewartii beta-carotene hydroxylase SEQ ID NO:12.

XX Pantoea stewartii; carotenoid biosynthetic enzyme; crtB; crtX; crtY;
XX crtI; crtZ; beta-carotene hydroxylase; enzyme; phytoene;
XX carotenoid.

XX Pantoea stewartii.

XX W02003016503-A2.

XX 27-FEB-2003.

XX 15-AUG-2002; 2002WO-US026647.

XX 15-AUG-2001; 2001US-0312646P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

XX WPI; 2003-268323/26.

XX N-PSDB; ACC44764.

XX Novel nucleic acid molecule isolated from Pantoea stewartii encoding a
XX carotenoid biosynthetic enzyme, useful for regulating carotenoid
XX biosynthesis in an organism.

XX Claim 4; Page 67-68; 68pp; English.

XX The present invention describes Pantoea stewartii carotenoid biosynthetic
XX enzymes (I). More specifically described are the geranylgeranyl
XX pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
XX lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
XX (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to
XX ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating
XX carotenoid biosynthesis in an organism, by over-expressing (I) in an
XX organism, such that the carotenoid biosynthesis is altered in the
XX organism. (I) and the genes encoding (I) are useful for converting
XX phytoene to the carotenoids, for creating recombinant organisms that have
XX the ability to produce various carotenoid compounds, and also for
XX enhancing or manipulating carotenoid compounds. (I) can also be used for
XX producing gene products having enhanced or altered activity

XX Sequence 175 AA;

Query Match 100.0%; Score 956; DB 6; Length 175;
Best Local Similarity 100.0%; Pred. No. 9.7e-108;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60

1 MLWLNALIVFTVVGMEVVAALAHKYIMHGWMGWSHLSHEPRKGAPEVNDLYAVVFAI 60
61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120
61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120
121 RMHHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGDGVDTSSTSGK 175
121 RMHHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGDGVDTSSTSGK 175

RESULT 4

LR07468
1 AAR07468 standard; protein; 175 AA.

2 AAR07468;

24-OCT-2003 (revised)

28-JAN-1991 (first entry)

Polypeptide with enzymatic activity for the conversion of beta-carotene into zeaxanthin.

Carotenoid biosynthesis; vitamin A; cancer; food coloring.

Pantoea ananatis.

EP333690-A.

24-OCT-1990.

20-APR-1990; 90BP-00107493.

21-APR-1989; 83JP-00103078.

05-MAR-1990; 90JP-00053255.

(KIRI) KIRIN BEER KK.

Misawa N, Kobayashi K, Nakamura K;

WPI; 1990-322212/43.

N-PSDB; AAQ06298.

DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.

Claim 6; Fig 6; 40pp; English.

Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

Sequence 175 AA;

Query Match 93.3%; Score 892; DB 2; Length 175;
Best Local Similarity 91.4%; Pred. No. 6e-100;
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

1 MLWLNALIVFTVVGMEVVAALAHKYIMHGWMGWSHLSHEPRKGAPEVNDLYAVVFAI 60

1 MLWLNALIVFTVVGMEVVAALAHKYIMHGWMGWSHLSHEPRKGAPEVNDLYAVVFAA 60

61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120

61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120

121 RMHHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGDGVDTSSTSGK 175

121 RMHHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGDGVDTSSTSGK 175

SULT 5

AAW87892
ID AAW87892 standard; protein; 175 AA.

XX AAW87892;

17-OCT-2003 (revised)

10-MAR-1999 (first entry)

Protein encoded by the carotenoid biosynthesis gene crtZ.

Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene; crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside; carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene; food additive.

Pantoea ananatis.

JP10327865-A.

15-DEC-1998.

29-MAY-1997; 97JP-00140460.

29-MAY-1997; 97JP-00140460.

(KIRI) KIRIN BREWERY KK.

(KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

WPI; 1999-099030/09.

N-PSDB; AAW84084.

New carotenoid glucoside(s) - used as food additives.

Disclosure; Page 23-24; 26pp; Japanese.

The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid glycoside, in which all, or part of, carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtZ, crtX or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 175 AA;

Query Match 93.3%; Score 892; DB 2; Length 175;
Best Local Similarity 91.4%; Pred. No. 6e-100;

Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLWLNALIVFTVVGMEVVAALAHKYIMHGWMGWSHLSHEPRKGAPEVNDLYAVVFAI 60

1 MLWLNALIVFTVVGMEVVAALAHKYIMHGWMGWSHLSHEPRKGAPEVNDLYAVVFAA 60

61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120

61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120

121 RMHHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGDGVDTSSTSGK 175

121 RMHHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQGDGVDTSSTSGK 175

RESULT 6

ABG93890

ID ABG93890 standard; protein; 175 AA.

XX ABG93890;

29-AUG-2003 (revised)

28-NOV-2002 (first entry)

X Erwinia uredovora beta-carotene hydroxylase.
 X Epsilon-cyclase; carotenoid; isopentenyl pyrophosphate isomerase;
 W beta-carotene hydroxylase; plant; enzyme.
 X X
 S Pantoea agglomerans.
 X US2002102631-A1.
 N X
 X 01-AUG-2002.
 D X
 X 02-JUN-1999; 99US-00323998.
 F X
 X 29-MAR-1996; 96US-00624125.
 R 25-SEP-1997; 97US-00937155.
 R 02-JUN-1998; 98US-00088724.
 R 02-JUN-1998; 98US-00088725.
 X X
 X (CUNN/) CUNNINGHAM F X.
 X (SUNZ/) SUN Z.
 X Cunnigham FX, Sun Z;
 X WPI; 2002-690609/74.
 X
 X Producing, or enhancing production of a carotenoid in a host cell by
 T transforming the cell with a vector encoding a epsilon cyclase enzyme.
 X
 X Example 1; Fig 6; 85pp; English.
 X
 X The invention relates to a method of producing or enhancing production of
 X a carotenoid in a host cell. The method comprises inserting into the cell
 X a vector comprising a heterologous nucleic acid sequence encoding a
 X protein with epsilon-cyclase enzyme activity operably linked to a
 X promoter, and expressing the heterologous sequence to produce the
 X protein. ABG93886-ABG93914 represent epsilon-cyclase, isopentenyl
 X pyrophosphate isomerase and beta-carotene hydroxylase proteins of the
 X invention. (Updated on 29-AUG-2003 to standardise OS field)
 X
 X Sequence 175 AA;
 SQ
 Query Match 90.0%; Score 860; DB 5; Length 175;
 Best Local Similarity 89.1%; Pred. No. 4.7e-96;
 Matches 156; Conservative 6; Mismatches 13; Indels 0; Gaps 0;
 QY 1 MLWLNALLIVFTVVGMEVVAALAHKYIMHGNGWHLSHHEPRKGAPEVNDLYAVVPAI 60
 DB 1 MLWLNALLIVFTVVGMEVVAALAHKYIMHGNGWHLSHHEPRKGAPEVNDLYAVVPAI 60
 QY 61 VSLIYFGSTGIMPLWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 120
 DB 61 LSLLIYFGSTGIMPLWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 120
 QY 121 RMHVAVRGEGCVSGFLYAPPLSKLOATLREHRAARSGAARDQDQGVDTSSSK 175
 DB 121 RMHVAVRGEGCVSGFLYAPPLSKLOATLREHRAARSGAARDQDQGVDTSSSK 175

OS Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39368).
 XX W09113078-A.
 XX 05-SEP-1991.
 XX 04-MAR-1991; 91WO-US001458.
 XX 02-MAR-1990; 90US-00487613.
 PR 18-MAY-1990; 90US-00525551.
 PR 03-AUG-1990; 90US-00562674.
 PR 28-FEB-1991; 91US-00662921.
 XX {STAD } AMOCO CORP.
 PA Ausich RL, Brinkhaus PL, Mukharji I, Proffitt JH, Yarger JG;
 PI Yen HC;
 XX WPI; 1991-281410/38.
 DR N-PSDB; AAQ13725.
 XX Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA
 PT encoding enzymes from Erwinia herbicola.
 XX Disclosure; Fig 21(1-3); 313pp; English.
 XX There are a total of six relevant genes in a 7900 bp region that cause E.
 CC coli cells to produce GGPP and the carotenoids phytoene through
 CC zeaxanthin diglucoside, which is the final prod. identified in the
 CC carotenoid pathway contd. in plasmid PARC376 (contg. a ca. 13 kb
 CC chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607
 CC (1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase,
 CC phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-
 CC carotene hydroxylase, and zeaxanthin glycosylase are represented in
 CC AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726
 CC respectively. The native sequence (AAQ13724) was genetically engineered.
 CC At the 5' end of the gene, codons encoding the native second and third
 CC amino acid have been changed from Leu-Val to Val-Leu. Recombinant
 CC expression plasmids can be used to produce large ants. of the enzymes and
 CC hence large amts. of the carotenoids which they synthesise. (Updated on
 CC 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-2003 to correct PR
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)
 XX Sequence 176 AA;
 SQ
 Query Match 68.4%; Score 654; DB 2; Length 176;
 Best Local Similarity 67.4%; Pred. No. 5.5e-71;
 Matches 116; Conservative 21; Mismatches 33; Indels 2; Gaps 1;
 QY 4 IWNALIVFTVVGMEVVAALAHKYIMHGNGWHLSHHEPRKGAPEVNDLYAVVPAI 63
 DB 2 VLNSLIIVLTIVLAMEGIAAFTTHYIMHGNGWHLSHHEPRKGVFELNDLFAVVPAI 61
 QY 64 ALIVFGSTGIMPLWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 123
 DB 62 ALIATGATAGVWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 121
 QY 124 HAVRGEKCVSGFLYAPPLSKLOATLREHRA--ARSGAARDQDQGVDTSSS 173
 DB 122 HAVRGEKCVSGFLYAPPLSKLOATLREHRA--ARSGAARDQDQGVDTSSS 173
 RESULT 8
 ABG93889
 ID ABG93889 standard; protein; 175 AA.
 XX
 AC ABG93889;
 XX
 DT 29-AUG-2003 (revised)
 DT 28-NOV-2002 (first entry)
 XX
 XX Erwinia herbicola beta-carotene hydroxylase.
 XX

A (KAIYO-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
 X WPI; 1999-09030/09.
 R N-PSDB; AAV84078.
 X T New carotenoid glucoside(s) - used as food additives.
 X S Disclosure; Page 14-15; 26pp; Japanese.
 X C The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid glycoside, in which all, or part of, carotenoid biosynthesis genes crtB, crtE, crtI, crtX, crtZ, or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives.
 X Q Sequence 162 AA;
 Query Match 52.2%; Score 499; DB 2; Length 162;
 Best Local Similarity 57.8%; Pred. No. 3.6e-52;
 Matches 93; Conservative 22; Mismatches 38; Indels 8; Gaps 4;
 Y 8 L I V F T V T V G M E V V A A L A H K Y I M H G - W G M G W H L S H H E P R K G A F E V N D I Y A V V P A I V S I A L I 66
 D 5 L I V A T V L V M E L T A Y S V H R W I M H G P L G M G W H K S H E H D H A L E K N D I Y G V F A V I A T V 64
 Y 67 Y F G S T G I M P - L W I G A G M T A Y G L L Y F M V H D G L V H Q R W P F R Y I P R K G Y L K R L Y M A H R M E H A 125
 D 65 T V G A Y W - N P V L M W I A L G M T V Y G L I F P V L H D G L V H Q R W P F R Y I P R G Y F R R L Y Q A H R L H E A 123
 Y 126 V R G K E C V S F G L Y A P P L S K L Q A T L R E H A A R S G A A R D E O D 166
 D 124 V E G R D H C V S F G I Y A P P V D K L K Q D L K - - - - - R S G V L R P Q D E 159

RESULT 11
 AAR79059
 ID AAR79059 standard; peptide; 162 AA.
 X X AAR79059;
 X X 27-AUG-2003 (revised)
 D T 28-FEB-1996 (first entry)
 X X
 X X 3 hydroxy-beta-ionone ring methylene to keto group converting peptide.
 X X Xanthophyll; astaxanthine; methylene; keto group; conversion;
 X X 3-hydroxy-beta-ionone ring.
 X X Agrobacterium aurantiacum.
 X X W09518220-A1.
 X X 06-JUL-1995.
 X X 26-DEC-1994; 94MO-JP002220.
 X X 27-DEC-1993; 93JP-00348737.
 X X 05-SEP-1994; 94JP-00235917.
 X X (KIRI) KIRIN BEER KK.
 X X (MARI-) MARINE BIOTECHNOLOGY INST CO LTD.
 X X Misawa N, Kondo K, Kajiwara S, Yokoyama A;
 X X WPI; 1995-246386/32.
 D R N-PSDB; AAG39486.
 X X DNA's encoding xanthophyll(s) - esp. asta:xanthin and other xanthophyll(s) using e.g. E. coli.

XX Claim 20; Page 79-80; 131pp; Japanese.
 PS
 XX AAR79058-R790629 are xanthophyll polypeptides. These polypeptides are capable of converting the 4-methylene group of a 3-hydroxy-beta-ionone ring to a 4-keto group in doing so these peptides also add a hydroxyl group to the 3-position carbon-atom of the 4-keto-beta-ionone ring. The DNA sequences may be used in the production of astaxanthine and other keto gp. cong. xanthophylls, the sequences may also be used to transform certain yeasts and other microorganisms. (Updated on 27-AUG-2003 to correct OS field.)
 X Q Sequence 162 AA;
 Query Match 51.8%; Score 495; DB 2; Length 162;
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;
 QY 6 N A L I V F T V T V G M E V V A A L A H K Y I M H G - W G M G W H L S H H E P R K G A F E V N D I Y A V V P A I V S I A 64
 D 3 N F L I V A T V L V M E L T A Y S V H R W I M H G P L G M G W H K S H E H D H A L E K N D I Y G L V F A V I A T V 62
 QY 65 L I Y F G S T G I M P - L W I G A G M T A Y G L L Y F M V H D G L V H Q R W P F R Y I P R K G Y L K R L Y M A H R M 122
 D 63 L F T V G - W I W A P V L M W I A L G M T V Y G L I F P V L H D G L V H Q R W P F R Y I P R K G Y A R R L Y Q A H R L 120
 QY 123 H H A V R G E G C V S F G L Y A P P L S K L Q A T L R E H A A R S G A A R D E 164
 D 121 H H A V E G R D H C V S F G I Y A P P V D K L K Q D L K - - - - - M S G V L R A E 157
 RESULT 12
 AAW82259
 ID AAW82259 standard; protein; 162 AA.
 X X
 X X AAW82259;
 X X 17-OCT-2003 (revised)
 D T 16-JUL-1999 (first entry)
 X X
 X X C. utilis crtZ protein.
 X X HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtZ;
 K W carotenoid.
 X X Pichia jadinii.
 X X JP10248575-A.
 X X 22-SEP-1998.
 X X 12-MAR-1997; 97JP-00058012.
 X X 12-MAR-1997; 97JP-00058012.
 X X (KIRI) KIRIN BREWERY KK.
 X X WPI; 1998-560727/48.
 D R N-PSDB; AAV73183.
 X X Gene useful for increase in carotenoid production - and preparation of carotenoid.
 X X
 X X Example 2; Fig 18-19; 54pp; Japanese.
 X X This invention describes a novel method for the preparation of carotenoids using genes and proteins isolated from Candida utilis. The invention specifically describes the isolation of a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence represents the Candida utilis crtZ protein which is used in the method of the invention. (Updated on 17-OCT-2003 to standardise OS field)
 X X Sequence 162 AA;
 PS

Query Match 51.8%; Score 495; DB 2; Length 162;
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64
 3 NFLIVVATVLMELTAYSVHRWIMHGPGWGWKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122
 63 LFTVG--WTAPVLMWIALGMTVYGLIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

123 HHAVRGKECVSGFGLYAPPLSKLOATLRRHAARSGAARDE 164
 121 HHAVERGRDHCVSFGFIYAPPVDKLKQDLK-----MSGVLRAE 157

RESULT 13
 AAW69536
 ID AAW69536 standard; protein; 162 AA.
 XX
 AC AAW69536;
 XX
 DT 10-AUG-1999 (first entry)
 XX
 DE crte2396 gene product beta-carotene hydroxylase.
 XX
 KW Carotenoid; pigment; canthaxanthin; R1534; crtB; prephytoene synthase;
 KW crtI; phytoene desaturase; crtY; lycopene cyclase; crtW2396;
 KW beta-carotene beta-oxygenase; food product; fermentation.
 XX
 OS Flavobacterium sp.
 XX
 FN JP10155497-A.
 XX
 PD 16-JUN-1998.
 XX
 PD 02-DEC-1997; 97JP-00348653.
 XX
 PD 02-DEC-1996; 96EP-00810839.
 XX
 PT (HOFF) HOPPMANN LA ROCHE & CO AG P.
 XX
 PS WPI; 1998-391048/34.
 XX
 PS N-PSDB; AAV40148.
 XX
 PT Preparation of carotenoid - comprises fermentation with transformed cell.
 XX
 PS Claim 2; Fig 64; 80pp; Japanese.
 XX
 CC The invention describes the preparation of carotenoid pigments e.g.
 CC canthaxanthins using a cell transformed by a vector having DNA sequences
 CC (a) to (e) or substantially homologous sequences. (a) a DNA sequence
 CC (crtE) coding GGGP synthase of Flavobacterium sp. R1534; (b) a DNA
 CC sequence (crtB) coding prephytoene synthase of Flavobacterium sp. R1534;
 CC (c) a DNA sequence (crtI) coding phytoene desaturase of Flavobacterium
 CC sp. R1534; (d) a DNA sequence (crtY) coding lycopene cyclase of
 CC Flavobacterium sp. R1534; and (e) a DNA sequence (crtW2396) coding beta-
 CC carotene beta-oxygenase of a microbe E-396 (FERM BP-4283). The carotenoid
 CC or a carotenoid mixture can also be used in preparation of food products.
 CC The method is an improved method of fermentation for carotenoid
 CC production
 XX
 SQ Sequence 162 AA;
 Query Match 51.8%; Score 495; DB 2; Length 162;
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64
 3 NFLIVVATVLMELTAYSVHRWIMHGPGWGWKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122
 63 LFTVG--WTAPVLMWIALGMTVYGLIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

123 HHAVRGKECVSGFGLYAPPLSKLOATLRRHAARSGAARDE 164
 121 HHAVERGRDHCVSFGFIYAPPVDKLKQDLK-----MSGVLRAE 157

QY 65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122
 DB 63 LFTVG--WTAPVLMWIALGMTVYGLIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

QY 123 HHAVRGKECVSGFGLYAPPLSKLOATLRRHAARSGAARDE 164
 DB 121 HHAVERGRDHCVSFGFIYAPPVDKLKQDLK-----MSGVLRAE 157

RESULT 14
 AAW87883
 ID AAW87883 standard; protein; 162 AA.
 XX
 AC AAW87883;
 XX
 DT 10-MAR-1999 (first entry)
 XX
 DE Protein encoded by the carotenoid biosynthesis gene crtZ.
 XX
 KW Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;
 KW crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;
 KW carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;
 KW food additive.
 XX
 OS Agrobacterium aurantiacum.
 XX
 FN JP10327865-A.
 XX
 PD 15-DEC-1998.
 XX
 PD 29-MAY-1997; 97JP-00140460.
 XX
 PD 29-MAY-1997; 97JP-00140460.
 XX
 PA (KIRI) KIRIN BREWERY KK.
 PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
 XX
 DR WPI; 1998-099030/09.
 DR N-PSDB; AAV84075.
 XX
 PT New carotenoid glucoside(s) - used as food additives.
 XX
 PS Disclosure; Page 11-12; 26pp; Japanese.
 XX
 CC The present sequence represents a protein involved in carotenoid
 CC biosynthesis. The specification describes astaxanthin diglucosides and
 CC adonixanthin-3'-glucosides. The specification also describes a method for
 CC the preparation of a carotenoid glycoside, in which all, or part of,
 CC carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtZ, crtX or crtW
 CC are introduced to a microbe or plant and expressed. The transformed
 CC organism is cultured and astaxanthin diglucosides, adonixanthin-3'-
 CC glucosides, and/or astaxanthin monoglucoside are collected. The
 CC carotenoid glucosides are used as food additives
 XX
 SQ Sequence 162 AA;
 Query Match 51.8%; Score 495; DB 2; Length 162;
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64
 3 NFLIVVATVLMELTAYSVHRWIMHGPGWGWKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122
 63 LFTVG--WTAPVLMWIALGMTVYGLIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

123 HHAVRGKECVSGFGLYAPPLSKLOATLRRHAARSGAARDE 164
 121 HHAVERGRDHCVSFGFIYAPPVDKLKQDLK-----MSGVLRAE 157

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:33:49 ; Search time 24.3391 Seconds
(without alignments)
3837.172 Million cell updates/sec

File: US-09-941-947a-34

rect score: 1535

quence: 1 MAVGSKSPATASTLPAKTR.....VTSRMKTYPPRAHLWQRP1 296

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	length	DB ID	Description
1	1535	100.0	296	2 Q8GCS0	Q8GCS0 pantoea ste
2	1371	89.3	296	2 Q47846	Q47846 pantoea agg
3	1244	81.0	296	2 Q8VUJ4	Q8VUJ4 pantoea agg
4	804	52.4	311	2 Q93C16	Q93C16 xanthobacte
5	743.5	48.4	335	2 Q9KTX1	Q9KTX1 bradyrhizob
6	686.5	44.7	303	2 P94789	P94789 flavobacter
7	645.5	42.1	304	2 Q9RLH3	Q9RLH3 paracoccus
8	446.5	29.1	346	2 Q8KZ27	Q8KZ27 uncultured
9	403	26.3	379	2 Q8RTY0	Q8RTY0 uncultured
10	395.5	25.8	344	2 Q9JPB3	Q9JPB3 rhodocyclu
11	362	23.6	325	16 Q9RW07	Q9RW07 geinococcus
12	351	22.9	343	2 O50567	O50567 rhodocyclu
13	313.5	20.4	422	10 Q9FV44	Q9FV44 tagetes ere
14	313.5	20.4	436	10 Q9XGW7	Q9XGW7 citrus para
15	310.5	20.2	303	16 Q7U419	Q7U419 synechococc
16	310.5	20.2	436	10 Q9M608	Q9M608 citrus unsh

17	309.5	20.2	399	10 Q8L8H7	Q8L8H7 tagetes ere
18	307.5	20.0	414	10 Q9PSY7	Q9PSY7 helianthus
19	302.5	19.7	414	10 Q9AVV8	Q9AVV8 helianthus
20	301.5	19.6	313	16 Q7V4E9	Q7V4E9 prochloroco
21	300	19.5	336	2 Q50892	Q50892 myxococcus
22	298	19.4	310	16 Q8KCM8	Q8KCM8 chlorobium
23	297.5	19.4	404	10 Q8W014	Q8W014 oryza sativ
24	296.5	19.3	272	10 Q8WVR1	Q8WVR1 oryza sativ
25	293.5	19.1	303	16 Q7VE48	Q7VE48 prochloroco
26	291.5	19.0	302	16 Q7V3D3	Q7V3D3 prochloroco
27	291	19.0	602	3 Q9P854	Q9P854 phycomyces
28	291	19.0	602	3 Q9P877	Q9P877 phycomyces
29	287.5	18.7	287	16 Q7NHD5	Q7NHD5 bordetella
30	286.5	18.7	287	16 Q7W9C9	Q7W9C9 bordetella
31	286.5	18.7	287	16 Q7VYT8	Q7VYT8 bordetella
32	286	18.6	432	10 Q40166	Q40166 lycopersico
33	284.5	18.5	389	10 Q9AU04	Q9AU04 haematococc
34	279.5	18.2	228	10 Q9LLR8	Q9LLR8 tagetes ere
35	276.5	18.0	425	10 Q04007	Q04007 dunaliella
36	267.5	17.4	311	16 Q8DIM2	Q8DIM2 synechococc
37	266	17.3	310	16 Q8YVY5	Q8YVY5 anabaena sp
38	261.5	17.0	290	16 Q9JRU9	Q9JRU9 neisseria m
39	255.5	16.6	230	16 Q9JUF5	Q9JUF5 neisseria m
40	245	16.0	331	16 Q9RIX8	Q9RIX8 streptomyce
41	240	15.6	221	10 Q84N52	Q84N52 zea mays (m
42	232	15.1	291	2 Q34288	Q34288 zymomonas m
43	232	15.1	342	16 Q93HP1	Q93HP1 streptomyce
44	231	15.0	153	10 Q84XR6	Q84XR6 citrus maxi
45	231	15.0	204	10 Q8L5H9	Q8L5H9 oryza sativ

ALIGNMENTS

RESULT 1

ID	Q8GCS0	PRELIMINARY;	PRT;	296 AA.
AC	Q8GCS0;			
DT	01-MAR-2003 (T-EMBLrel. 23, Created)			
DT	01-MAR-2003 (T-EMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)			
DB	PHYTOENE SYNTHASE.			
GN	CRFB.			
OS	Pantoea stewartii.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoea.			
OX	NCBI_TaxID=56269;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 8200;			
RA	deSouza M.L., Kollmann S.R., Schroeder W.A.;			
RT	"Carotenoid Biosynthesis (WO 02/079395 A2).";			
RL	Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AV166713; AAN85600.1; -			
DR	GO; GO:0016740; F:transferase activity; IEA.			
DR	GO; GO:0009058; P:biosynthesis; IEA.			
DR	InterPro; IPR002060; Squ/phyt_synthase.			
DR	InterPro; IPR008949; Terpenoid_synthase.			
DR	Pham; PF00494; SQS_PSY; 1.			
DR	PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.			
DR	PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.			
SQ	SEQUENCE 296 AA; 33108 MW; B2294B16D8513PC2 CRC64;			

Query Match 100.0%; Score 1535; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 4.2e-126;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAVGSKSPATASTLPAKTRSVLMYAMCRCHDDVIDDQTLGFGHADQPSQMPERLQ 60
Db	1	MAVGSKSPATASTLPAKTRSVLMYAMCRCHDDVIDDQTLGFGHADQPSQMPERLQ 60
Qy	61	LEMKTRQAVAGSOMPEPAFAFQEVAMAHDIAPAYAFHLEGFAMDVRETRYLTLDLTLR 120

```
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHEGFAMDVRETRYLTLDLTLR 120
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDGLAQFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDGLAQFQLTNIARDIVDDAQVGRCYLPESWL 180
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240
241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 2
47846 PRELIMINARY; PRT; 296 AA.
C Q47846;
T 01-NOV-1996 (TREMELrel. 01, Created)
T 01-NOV-1996 (TREMELrel. 01, Last sequence update)
T 01-OCT-2003 (TREMELrel. 25, Last annotation update)
E Prephosphatase synthase.
C CRTC.
S Pantoea agglomerans.
S Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=549;

Query Match 89.3%; Score 1371; DB 2; Length 296;
Best Local Similarity 88.5%; Pred. No. 9.5e-112;
Matches 262; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Y 1 MAVGSKSPATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSOMPQRLOQ 60
1 MAVGSKSPATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSOMPQRLOQ 60
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHEGFAMDVRETRYLTLDLTLR 120
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHEGFAMDVRETRYLTLDLTLR 120
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDGLAQFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDGLAQFQLTNIARDIVDDAQVGRCYLPESWL 180
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240
241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 3
```

```
Q8VUJ4 PRELIMINARY; PRT; 296 AA.
ID Q8VUJ4;
AC Q8VUJ4;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE CrtB protein.
GN CRTC.
OS Pantoea agglomerans pv. milletiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=182454;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamiyama H., Hirata R.;
RT "Isolation and characterization of carotenoid biosynthesis genes from
RT Pantoea agglomerans pv. milletiae Wist 801."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076662; BAB79604.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR002060; Squ/phyt_synthase.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00494; SQS_PSY; 1.
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
SQ SEQUENCE 296 AA; 32696 MW; C2BFC3E0C3C3CE56 CRC64;

Query Match 81.0%; Score 1244; DB 2; Length 296;
Best Local Similarity 82.1%; Pred. No. 1.2e-100;
Matches 243; Conservative 12; Mismatches 41; Indels 0; Gaps 0;

QY 1 MAVGSKSPATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSOMPQRLOQ 60
DB 1 MEVGSKSPATSKLFGAKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSOMPQRLOQ 60
QY 61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHEGFAMDVRETRYLTLDLTLR 120
DB 61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHEGFAMDVRETRYLTLDLTLR 120
QY 121 YCHVAGVGLMMAQIMGVDRDNATLDRACDGLAQFQLTNIARDIVDDAQVGRCYLPESWL 180
DB 121 YCHVAGVGLMMAQIMGVDRDNATLDRACDGLAQFQLTNIARDIVDDAQVGRCYLPESWL 180
QY 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240
DB 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240
QY 241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
DB 241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 4
Q93C16 PRELIMINARY; PRT; 311 AA.
ID Q93C16;
AC Q93C16;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Phytoene synthase.
GN CRTC.
OS Xanthobacter sp. (strain Py2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Xanthobacter.
OX NCBI_TaxID=78245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Py2;
RA Larsen R.A., Metcalf W.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF408848; AAL02001.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
```


GO; GO:0009058; P:biosynthesis; IEA.
InterPro; IPR002060; Squ/phyt synthase.
InterPro; IPR008949; Terpenoid synthase.
Pfam; PF00494; SOS_PSY; 1.
PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
SEQUENCE 311 AA; 33279 MW; 7752769786099CAA CRC64;
Query Match 52.4%; Score 804; DB 2; Length 311;
Best Local Similarity 54.9%; Pred. No. 4.4e-62;
Matches 161; Conservative 37; Mismatches 93; Indels 2; Gaps 2;
4 GSKSPATASTLFDKATRSVLMYAWCRHCDVDVDDTGLGFHADQPSQMPERLQOLEM 63
19 GSKSPAAARLFGPRMEDAYMLYAWCRHCDVDVDDGQELG-HGVSVAASPAERLDGLYE 77
64 KTRQAYAGSQHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYTLTLDOTLRYCY 123
78 ETRNAYRGAPSAHPAFAPFAEVVKNDEIPERYPLQLLEGFMVDEGGRYDTLDTLTYCY 137
124 HVAGVGLMAYQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEEE 183
138 HVAGVGVVYXALIMGARDEVVLDASDLGLGFQLTNIARDVMDAALIGRIYLPFAFLAEA 197
184 GLTKANYAAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYRKIG 243
198 GVPERAIAAPERRAAVAGVVARLLDVAEPPYDQALIGMAALPFAAARAAVGAARGVYRAIG 257
244 VKVQAGQKQWHDHROSTSTABKLLTLLTASQAVTSRMK-TYPERPAHLWQRP 295
258 IEVRRKRGKAWDTRVSTSTAQAGFLAGGLAQAALATFRGAPPRPAHLWTRP 310
SULT 5
KIXI
Q9KIX1 PRELIMINARY; PRT; 335 AA.
01-OCT-2000 (Tremblrel. 15, Created)
01-OCT-2000 (Tremblrel. 15, Last sequence update)
01-OCT-2003 (Tremblrel. 25, Last annotation update)
Phytoene synthase.
CRTB.
Bradyrhizobium sp. ORS278.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
NCBI_TaxID=114615;
[1]
SEQUENCE FROM N.A.
STRAIN=ORS278;
MEDLINE=20309720; PubMed=10851005;
Hannibal L., Lorquin J., Argles d'Ortoli N., Garcia N.,
Chaintreuil C., Masson-Bovin C., Dreyfus B., Giraud E.,
"Isolation and characterization of the canthaxanthin biosynthesis
genes from the photosynthetic bacterium Bradyrhizobium sp. strain
ORS278";
J. Bacteriol. 182:3850-3853 (2000).
EMBL; AF218415; AAF78202.1; -
GO; GO:0016740; P:transferase activity; IEA.
GO; GO:0009058; P:biosynthesis; IEA.
InterPro; IPR002060; Squ/phyt synthase.
InterPro; IPR008949; Terpenoid synthase.
Pfam; PF00494; SOS_PSY; 1.
PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
SEQUENCE 335 AA; 36869 MW; 6010A5B12014FDD2 CRC64;
Query Match 48.4%; Score 743.5; DB 2; Length 335;
Best Local Similarity 51.7%; Pred. No. 9.6e-57;
Matches 156; Conservative 40; Mismatches 87; Indels 19; Gaps 4;

QY 62 EMKTRQAYAGSQHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYTLTLDTLRY 121
DB 78 RDQTAQALEGAPMEDPVFGIQRVVQHEAIPHHVFEVLLDGFAMDVGREYETLSETLDY 137
QY 122 CYHVAGVGLMAYQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLE 181
DB 138 CYHVAGVGVVYXALIMGARDEVVLDASDLGLGFQLTNIARDVMDAALIGRIYLPFAFL 197
QY 182 BEGLTKANYAAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYRK 241
DB 198 EAGVPAEAEVAPERRAAVAGVVARLLDVAEPPYDQALIGMAALPFAAARAAVGAARGVY 257
QY 242 IGKVEQAGKQWHDHROSTSTAEK-----LTLTASQAVTSRMKTYPPPAHLWQ 293
DB 258 IGRVKKRGKAWDTRVSTSTAQAGFLAGGLAQAALATFRGAPPRPAHLWTRP 310
QY 294 RP 295
DB 311 RP 312
RESULT 6
P94789 PRELIMINARY; PRT; 303 AA.
AC P94789
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Phytoene synthase.
GN CRTB.
OS Flavobacterium sp. ATCC 21588.
OC Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales;
OC Flavobacteriaceae; Flavobacterium.
OX NCBI_TaxID=50286;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=R1534;
RX MEDLINE=97186694; PubMed=9034310;
RA Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,
van Loon A.P.;
RT "Isolation and characterization of the carotenoid biosynthesis genes
of Flavobacterium sp. strain R1534";
RL Gene 185:35-41 (1997).
DR EMBL; U62808; AAC44849.1; -
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR002060; Squ/phyt synthase.
DR InterPro; IPR008949; Terpenoid synthase.
DR Pfam; PF00494; SOS_PSY; 1.
DR PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
DR PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
SQ SEQUENCE 303 AA; 32616 MW; 536CC680B4D9D0 CRC64;
Query Match 44.7%; Score 686.5; DB 2; Length 303;
Best Local Similarity 48.7%; Pred. No. 8.2e-52;
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;
QY 1 MAVSKSPATASTLFDKATRSVLMYAWCRHCDVDVDDTGLGFHADQPSQMPERLQ 60
DB 12 IAQGSQSPAQAALMPGIREDTVMYAWCRHADDVIDGQVMSAPEAGGD--PQALSLGA 69
QY 61 LEMKTRQAYAGSQHE-----PAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYTL 115
DB 70 LEADTLAA-----LHEDGPMSPPPAALRQVARRHDFPDLMPMDLIEGFAMDVARRRSL 124
QY 116 DDTLRYCYHVAGVGLMAYQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYL 175
DB 125 DDVLEYSYHVAGVGVVYXALIMGARDEVVLDASDLGLGFQLTNIARDVMDAALIGRIY 184
QY 176 PESWLEEGTKANYAAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATA 235
DB 185 PADWLAEG---ATVEGVPFSDALYSVIRLLDAAEPPYASARQGLPHLPPCAWSIAAA 241

236 KQYRKIGVQVQAGQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 295
 242 LRIYRAIGTWIRGGPEAYRQISTSKAKIGLLARGGLDAAASRLRGHISRDGLWTRP 301

RESULT 7

9RLH3 D Q9RLH3 PRELIMINARY; PRT; 304 AA.
 C Q9RLH3; 01-MAY-2000 (TrEMBLrel. 13, Created)
 T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 E Phytoene synthase.
 N CRFB.
 S Paracoccus marcusii.
 C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 C Rhodobacteraceae; Paracoccus.
 X NCBI_TaxID=59779;
 P [1]
 N SEQUENCE FROM N.A.
 C STRAIN=MHI;
 A Harker M., Hirschberg J.;
 T "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
 T MHI";
 L Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 R ENBL; Y15112; C856063.1; --
 R GO; GO:0016740; P:transferase activity; IEA.
 R GO; GO:0009058; P:biosynthesis; IEA.
 R InterPro; IPR002060; Squ/phyt_synthase.
 R InterPro; IPR008949; Terpenoid_synth.
 R Pfam; PF00494; SQS_PSV; 1.
 Q SEQUENCE 304 AA; 33185 MW; 0C0BA7CEDC30828A CRC64;

Query Match 42.1%; Score 645.5; DB 2; Length 304;
 Best Local Similarity 46.9%; Pred. No. 3.2e-48;
 Matches 137; Conservative 35; Mismatches 115; Indels 5; Gaps 2;

Y 4 GSKSFATATLFDKTRRSVLMYAWCRHCDVDIDDTLGFHADQPSQMPERLQOLEM 63
 b 15 GSQSFATAXLMPGGIRDTVWLYAWCRHADDVIOGLAGSRPEAVND--FQARLDGLRA 72
 Y 64 KTRQAYAGSQMHEPAPAFQEVAMADHIAFADPHLEGFAMDVRETRVLTLDLDTLRYCY 123
 b 73 DTLAALQGGPGVTPPPFAALRAVARRHDPQAWPMDLIEGFAIDVEARDYRTLDVLEYSY 132
 Y 124 HVAGVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEEE 183
 b 133 HVAGILGVMMARVGVDRHVPVLDRACDLGLAFQLTNIARDIVDYGRCYLPGLDHLA 192
 Y 184 GLTKANYAAPENRQALSRIAGRLVREASPYVYSSMAGLAQLPLRSAMAIATAKOVYRKIG 243
 b 193 G---ARVDGVPVSPSLYTVILDLDAEFPYPSARVGLADUPPCAWSIPAAFRIVRAIG 249
 Y 244 VKVQAGQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 295
 b 250 LPIRKGPEAYRQISTSKAKIGLLGIGGWMDVADHACRGSGVSRQDLWTRP 301

RESULT 8

3KZ27 D Q8KZ27 PRELIMINARY; PRT; 346 AA.
 C Q8KZ27; 01-OCT-2002 (TrEMBLrel. 22, Created)
 T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 E Phytoene synthase.
 N CRFB.
 S uncultured proteobacterium.
 C Bacteria; Proteobacteria; environmental samples.
 X NCBI_TaxID=153809;
 P [1]
 N SEQUENCE FROM N.A.

MEDLINE=21822632; PubMed=11832943;
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
 RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;
 RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
 RL Nature 415:630-633(2002).
 DR EXBL; AE008921; AAM48647.1; --
 DR GO; GO:0016740; P:transferase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR InterPro; IPR002060; Squ/phyt_synthase.
 DR InterPro; IPR008949; Terpenoid_synth.
 DR Pfam; PF00494; SQS_PSV; 1.
 DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
 DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
 SQ SEQUENCE 346 AA; 37857 MW; 13A5C6B82FAC1237 CRC64;

Query Match 29.1%; Score 446.5; DB 2; Length 346;

Best Local Similarity 38.0%; Pred. No. 1e-30;
 Matches 115; Conservative 35; Mismatches 120; Indels 33; Gaps 5;

QY 4 GSKSFATATLFDKTRRSVLMYAWCRHCDVDIDDTLGFHADQPSQMPERLQOLEM 63
 Db 17 GSLSFHAASKLLPASVRDPAALAYAFCLADDEVDE-----GQNKTRAVIELQE 65
 QY 64 KTRQAYAGSQMHEPAPAFQEVAMADHIAFADPHLEGFAMDVRETRVLTLDLDTLRYCY 123
 Db 66 RLALVYAGPRNAPEDRAPASVVEDFEMPAALPEALLEGLANDEHRYSSLSDLRGYCA 125
 QY 124 HVAGVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEEE 183
 Db 126 RVASAGVAMCVMVRVDADALARACDLGVAMQLTNIARDIVGEDARAGRIVLEWIDAE 185
 QY 184 GL-----TKANYAAPENRQALSRIAGRLVREASPYVYSSMAGLAQLPLRSAMAIATAKOV 238
 Db 186 GLDPQVLSVTEATPE-----LRMWKLLSEHAALYVRSEAGVAALPLNARTGIVYARIY 241
 QY 239 YRKIGVQVQAGQAWDHROSTSTAEKLTLL-----TASGQAVTSRMKTYPPRPAHLWQ 293
 Db 242 YDAIGQAVARNHYDSITHRGRTTKAKMALLAKSLRTAAGLV-----PRSPVLVA 293
 QY 294 RPI 296
 Db 294 RPL 296

RESULT 9

Q8RTY0 ID Q8RTY0 PRELIMINARY; PRT; 379 AA.
 AC Q8RTY0; 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Phytoene synthase.
 GN CRFB.
 OS uncultured proteobacterium.
 OC Bacteria; Proteobacteria; environmental samples.
 OX NCBI_TaxID=153809;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21822632; PubMed=11832943;
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
 RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;
 RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
 RL Nature 415:630-633(2002).
 DR EXBL; AE008919; AAL76346.1; --
 DR GO; GO:0016740; P:transferase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR InterPro; IPR002060; Squ/phyt_synthase.
 DR InterPro; IPR008949; Terpenoid_synth.
 DR Pfam; PF00494; SQS_PSV; 1.
 DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
 SQ SEQUENCE 379 AA; 41383 MW; 3AAL605561C3CC6D CRC64;

Query Match 26.3%; Score 403; DB 2; Length 379;

Best Local Similarity 38.4%; Pred. No. 7.3e-27;
Matches 103; Conservative 38; Mismatches 115; Indels 12; Gaps 4;

4 GSKSFATSLTDFAKTRSVLMLYAWCRHCDVDVDTQTLGFHADPSSOMPQRLOQLQM 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 GSRSEFWASQLLPPTLRNDACGLYAFCREADDLLDE--GDDAALAQLHE-RLDGI-- 94
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

64 KTRQAYAGSQMHEPAFAFOEVAMAHIDIAIPAYAPDHLGEFAMDRETRYILDDTLRYCY 123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 -----YSGAQPQERTVDRVLQRIVTHHQPLRTLLALEGCFAMNDASGRHYHTLSDPAYGA 149
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

124 HVAGVVGLMMAOIMGVRONATILDRACDLGLAFOLNTIARDIVDDAQVGRCYLPSWLEE 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

150 RVAGVVGVMAVLMGVRSNALARADLVAMQLTIARDVGEDARGLYLFPRQMLUDA 209
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

194 GLTKANY-AAENRQALSRIAGRLVRERAEFYVSSMAGLAQLPLRSANAIATAKOVYRKI 242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

210 GDIPSFLQAFTFSPALCTVVERILLLEAEOLYRSSESIGALLPVGARPCIVAARLLVAEI 269
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

243 GVKVROAGKGQAMDEHQSTSTAETKLTLII 270
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

270 GRALLKSGNSIDIRAYIGMAGKARLLVL 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
JJPB3

Q3JPE3 PRELIMINARY; PRT; 344 AA.

01-OCT-2000 (TREMBLrel. 15, Created)
01-OCT-2000 (TREMBLrel. 15, Last sequence update)
01-OCT-2003 (TREMBLrel. 25, Last annotation update)
Phosphotene pyrophosphate synthase.
CPTB.

Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Comamonadaceae; Rubrivivax.
NCBI_TaxId=288668;
[1]

SEQUENCE FROM N.A.
STRAIN=IL144;
Nagashima K.V., Shimada K., Matsura K.;
"Phylogenetic analysis of photosynthetic genes of Rhodocyclus
gelatinosus: Possibility of horizontal gene transfer in purple
bacteria";
Photosyn. Res. 36:195-191(1993).
[2]

SEQUENCE FROM N.A.
STRAIN=IL144;
MEDLINE=94132007; PubMed=8300574;
Nagashima K.V., Matsura K., Ohya S., Shimada K.;
"Primary structure and transcription of genes encoding B870 and
photosynthetic reaction center apoproteins from Rubrivivax
gelatinosus,";
J. Biol. Chem. 269:2477-2484(1994).
[3]

SEQUENCE FROM N.A.
STRAIN=IL144;
Igarashi N., Shimada K., Matsura K., Nagashima K.V.;
"Photosynthetic gene cluster in purple bacterium, Rubrivivax
gelatinosus,";
(in) Garab G. (eds.);
Photosynthesis:
mechanisms and effects (Proceedings of the 11th international congress
on photosynthesis), pp.4:2889-2892, Kluwer Academic Publishers,
Dordrecht (1999).
[4]

SEQUENCE FROM N.A.
STRAIN=IL144;
MEDLINE=20031519; PubMed=10563807;
Marion L., Yoshida M., Jaquind M., Nagashima K.V., Matsura K.,
Farot P., Vermeglio A.;
"Dark aerobic growth conditions induce the synthesis of a high
midpoint potential cytochrome c8 in the photosynthetic bacterium
Rhodospirillum rubrum";
Biochim. Biophys. Acta 1459:105-114(1999).
[5]

```

RT Rubrivivax gelatinosus.";
RL Biochemistry 38:15238-15244 (1999).
EMBL AB034704; BAA94048.1; -.
PIR T50895; T50895.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0009058; P:biosynthesis; IEA.
DR InterPro: IPR002060; Squ/phyt_synthese.
DR IntraPro: IPR008949; Terpenoid_synth.
DR Pfam: PF00494; SQS_PSY; 1.
DR PROSITE: PS01045; SQUALEN PHYTOEN SYN 2; 1.
SQ SEQUENCE 344 AA; 37551 MW; CCG7F8AD52AA4570 CRC64;

Query Watch 25.8%; Score 395.5; DB 2; Length 344;
Best Local Similarity 36.2%; Pred.No. 2.9e-26;
Matches 106; Conservative 32; Mismatches 140; Indels 15; Gaps 3;

QY 1 MAVGSKSFATASTIFDAKTSSVLMUYAMCRHCDVDIDDTLGFHADQPSOMPQRLEQQ 60
DB 18 MRGGSKSPFAASLLLPQVRTPATALYACRVADDADV-----LSGDPHAAMDE 66
QY 61 LEMKTRQYAGSQMHSEPAFAAFQEVMAHDIPAPAFDHLEGFMADVRETRYLIUDTLR 120
DB 67 LRTRLDVAVYAGTAPIAADRALASTVHRYGVPVRLLDALLEGFLWDADGRDYTDVVEA 126
QY 121 YCYHVAGVGLMWAQINGVRDNATLRDCDLGLAFOLNTIARDIVDDAQVGRCYLPESWL 180
DB 127 YCARVAGTVGAMALINGVASPQALARACELGVAMQFTNIARDVCEADANGELYLPREW 186
QY 181 EEEGL-TKYANYAAPFNQALRSIAGRLVREASEPYVSVSMAGLAQLPLRSAWAIATAKVY 239
DB 187 VEAGLEDVDAMLQNPHVCVAQVTQVRELLRAADELYERSEHGIIAALPRCPAIRAARLV 246
QY 240 RKGIVGYEQACKQANDHQSTSTAETKLITLLTASGOAVTSRMKTY---PPRPA 289
DB 247 AEIGMLERQGLDSVNRHVVPARRKAALMARASAANFPNGRAYISMPPLPA 299

RESULT 11
QSRW07 PRELIMINARY; PRT; 325 AA.
ID QSRW07 PRELIMINARY; PRT; 325 AA.
AC QSRW07.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phytoene synthase.
DE DE
GN DR0862.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcaceae;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
[1] RN
RN RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.P., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Hatt D.H., Quinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Yamahavan J.S., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.V., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
EMBL AB001940; AAF10440.1; -.
DR PIR; D75466; D75466.
DR TIGR; DR0862; -.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0009058; P:biosynthesis; IEA.
DR InterPro: IPR002060; Squ/phyt_synthese.
DR IntraPro: IPR008949; Terpenoid_synth.
DR Pfam: PF00494; SQS_PSY; 1.
DR PROSITE: PS01045; SQUALEN PHYTOEN SYN 2; 1.

```


Citrus paradisi (Grapefruit).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Sapindales; Rutaceae; Citrus.
NCBI_TaxID=37656;
[1]
SEQUENCE FROM N.A.
Costa M.C., Moreira C.D., Melton J.R., Otoni W.C., Moore G.A.;
"Developmental expression of carotenoid genes in Citrus";
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF152892; AAD38051.2; ..
GO: 0016740; P:transferase activity; IEA.
GO: 0009058; P:biosynthesis; IEA.
InterPro; IPR002060; Squ/phyt synth.
Pfam; PF00494; SOS_PSY; 1.
PROSITE; PS01044; SQUALEN_PHYTOEN SYN 1; 1.
PROSITE; PS01045; SQUALEN_PHYTOEN SYN 2; 1.
SEQUENCE 436 AA; 49360 MW; 87F7C8D798FA8B1 CRC64;
Query Match 20.4%; Score 313.5; DB 10; Length 436;
Best Local Similarity 28.1%; Pred. No. 6.1e-19;
Matches 81; Conservative 56; Mismatches 132; Indels 19; Gaps 5;
5 SKSPATASTLPDAKTRRSVLMYAKCRHCDVIDDQTLGFHADQPSOMPQRLOQLEMK 64
156 AKTFYLGTLMTSERRALWAIYVWCRTDELVDGPN-----ASHITPDLRWESR 207
65 TQAVAGSQMHPEAPAFQEVANAHK--DIAPAYAFDHLGFMADVRETRYLTDDTLRYC 122
208 LEDLFRG-QPFDLDAALSDTVTKFPVDIQPFR--DMTEGVRMDLRKSRKYNFDELYLC 264
123 YHVAGVGLMAQIMGVDRN-----ATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
265 YVAGTVGLMSVPGIAPDSQATTESVYNALAGIANQLTNILRDVGDGRPRGYLP 324
177 ESWLEEEGLTKANYAAPENRQALSRIAGRLVREAPYVYSSMAGLAQLPLSAWAIATAK 236
325 QDLAQAGLSDDIDPAGEVTIKWRNFKNQIKRSMFFDMAENGVTLSSEASRWPVWASL 384
237 QYRKIGVYKVGQAGQKQMDHROSTSTAEKLTLLLTASQAVTSRMKTY 284
385 LLYRQILDIEANDYNNFTKACVSKAKXKXIALPIAYAKSLLRPSRIY 432
SEQUENCE FROM N.A.
MEDLINE=2825697; PubMed=12917641;
Palenik B., Brahansha B., Larimer F.W., Land M., Hauser L., Chain P.,
Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,
Bufreres A., Partensky F., Webb E.A., Waterbury J.;
"The genome of a motile marine Synechococcus";
Nature 424:1037-1042(2003).
EMBL; BX569695; CAE08771.1; ..
Transferase; Complete proteome.
SEQUENCE 303 AA; 34984 MW; 05527FA419F80005 CRC64;
Query Match 20.2%; Score 310.5; DB 16; Length 303;
Best Local Similarity 31.8%; Pred. No. 6.8e-19;
Matches 87; Conservative 43; Mismatches 125; Indels 19; Gaps 6;

QY 5 SKSPATASTLPDAKTRRSVLMYAKCRHCDVIDDQTLGFHADQPSOMPQRLOQLEMK 64
DB 24 AKTFYLGTLMTSERRALWAIYVWCRTDELVDGPN-----ASHITPDLRWESR 79
QY 65 TQAVAGSQMHPEAPAFQEVANAHK--DIAPAYAFDHLGFMADVRETRYLTDDTLRYC 123
DB 80 TQALFNGTVENDLDAMVDTLTERFPQDIQ--YLDMEGQRMDLTWIRYPRFDDLKLYCY 137
QY 124 YHVAGVGLMAQIMGVDRN-----ATLDRACDLGLAFQLTNIARDIVDDAQVGR 172
DB 138 RVAGTVGLMTQGMGV-DGAVTSAPWSRDPDTSAAVALGIANQLTNILRDVGDGRGR 196
QY 173 CYLPSWLEEEGLTKANYAAPENRQALSRIAGRLVREAPYVYSSMAGLAQLPLSAWAI 232
DB 197 IYLPLEDLERFCYSDDLLMAGRLNSAWCELQFQKRDARDFARSEAGVRWLSRDARW 256
QY 233 ATAKQVYRKIGVYKVGQAGQKQMDHROSTSTAEKLTLLLTASQAVTSRMKTY 266
DB 257 WTSRLRYRGILDIERVDYDVFNNRYVGVKYSKL 290

Search completed: February 29, 2004, 14:51:04
Job time : 27.3391 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 4.98876 Seconds
(without alignments)
3089.496 Million cell updates/sec

title: US-09-941-947A-34

effect score: 1535

sequence: 1 NAVGSKSPATSTLPDANKT.....VTSRMKTYPPRAHLWQRPPI 296

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1368	89.1	309	1 CRTB_PANAN	P21683 pantoeca ana
2	1028	67.0	309	1 CRTB_BRWHE	P22872 erwinia her
3	648	42.2	301	1 CRTB_AGRAU	P54975 agrobacteri
4	394	25.7	355	1 CRTB_RHOSH	P54905 rhodobacter
5	393.5	25.6	339	1 CRTB_RHOCA	P17056 rhodobacter
6	328.5	21.4	423	1 PSY_NARPS	P53797 nartissus p
7	309.5	20.2	419	1 PSY_CAPAN	P37272 capsicum an
8	307.5	20.0	310	1 PSY2_LYCES	P37273 lycopersico
9	307.5	20.0	398	1 PSY_DAUCA	Q98808 daucus caro
10	307.5	20.0	422	1 PSY_ARATH	P37271 arabidopsis
11	302.5	19.7	410	1 PSY_MAIZE	P49085 zea mays [m
12	300.5	19.6	412	1 PSY1_LYCES	P08196 lycopersico
13	297	19.3	602	1 PSY_NEUCR	P37295 neurospora
14	286	18.5	422	1 PSY_CUCME	P49293 cucumis mel
15	280	18.2	309	1 CRTY_SPIPL	O07333 spirulina p
16	273.5	17.8	614	1 LCPS_RHIRA	Q94066 rhizomucor
17	273	17.8	289	1 CRTB_THERH	P37270 thermus the
18	265.5	17.3	337	1 CRTB_SNNY3	P37294 synechocyst
19	258	16.8	307	1 CRTB_SNNP7	P37269 synechococc
20	233.5	15.2	342	1 CRTB_STRGR	P54977 streptomyce
21	217.5	14.2	302	1 CRTB_WYCTU	O50728 mycobacteri
22	207	13.5	279	1 Y4AC_RHISN	P53350 rhizobium s
23	171.5	11.2	417	1 FDFT_HUMAN	P37268 homo sapien
24	163.5	10.7	416	1 FDFT_MOUSE	P37398 mus musculu
25	159.5	10.4	416	1 FDFT_RAT	Q02769 rattus norv
26	159	10.4	292	1 Y4AD_RHISN	P53351 rhizobium s
27	149	9.7	443	1 FDFT_CANGA	Q98626 candida gla
28	145.5	9.5	448	1 FDFT_CANAL	P78589 candida alb
29	130.5	8.5	460	1 FDFT_SCHPO	P36596 schizosacch
30	124	8.1	444	1 FDFT_YEAST	P29704 nicotiana b
31	116.5	7.6	411	1 FDFT_NICEE	P53800 nicotiana b
32	116	7.6	410	1 FDFT_ARATH	P53799 arabidopsis
33	109.5	7.1	572	1 FDFT_USTMA	Q92459 ustilago ma

ALIGNMENTS

RESULT 1

ID	CRTB_PANAN	STANDARD;	PRT;	309 AA.
AC	P21683	1991 (Rel. 18, Created)		
DT	01-MAY-1991	(Rel. 42, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Phytoene synthase (EC 2.5.1.-)			
GN	CRTB.			
OS	Pantoeca ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoeca.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
EX	MEDLINE=91072214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway			
RT	by functional analysis of gene products expressed in Escherichia			
RT	coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
RN	[2]			
RP	REVISIONS TO N-TERMINUS.			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.			
CC	-I- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to			
CC	phytoene.			
CC	-I- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +			
CC	prephytoene diphosphate.			
CC	-I- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +			
CC	phytoene.			
CC	-I- PATHWAY: Carotenoid biosynthesis.			
CC	-I- SIMILARITY: Belongs to the phytoene/squalene synthetase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; D90087; BAA14128.2; --			
DR	PIR; E37802; E37802.			
DR	InterPro; IPR002060; Squ/phyt_synth.			
DR	InterPro; IPR008949; Terpenoid_synth.			
DR	Pfam; PF00494; SQS_PSY; 1.			
DR	PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.			
DR	PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.			
KW	Multifunctional enzyme; Carotenoid biosynthesis; Transferase.			
SQ	SEQUENCE 309 AA; 34472 MW; 9AA381A7376BBFC9 CRC64;			
Query Match	89.1%; Score 1368; DB 1; Length 309;			

Best Local Similarity 88.9%; Pred. No. 5.6e-115;
Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

1 MAVGSKSPATASTLFAKTRRSVLMYAWCRHCDVDVDDQTLGFGFADQSSQWPERLQO 60
14 MAVGSKSPATASTLFAKTRRSVLMYAWCRHCDVDVDDQTLGFGFADQSSQWPERLQO 73
61 LEMKTRQAYAGSQHHEPAPAFQEVAMAHDIAPAYAFDHLGFGFADQSSQWPERLQO 120
74 LEMKTRQAYAGSQHHEPAPAFQEVAMAHDIAPAYAFDHLGFGFADQSSQWPERLQO 133
121 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQCRCYLPESWL 180
134 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQCRCYLPESWL 193
181 EESGLTKANYAAPENRQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 240
194 EHEGLNKENTAAPENRQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 253
241 KIGVKVEQAGQAWDRQSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
254 KIGVKVEQAGQAWDRQSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 309

RESULT 2
RTB_ERWHE STANDARD; PRT; 309 AA.
C P22872, 1991 (Rel. 19, Created)
01-AUG-1991 (Rel. 19, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene synthase (EC 2.5.1.-).
Erwinia herbicola.
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Pantoea.
NCBI_TaxID=549;
SEQUENCE FROM N.A.
STRAIN=EH010;
MEDLINE=91088634; PubMed=2263648;
Armstrong G.A., Alberti M., Hearst J.E.;
"Conserved enzymes mediate the early reactions of carotenoid
biosynthesis in nonphotosynthetic and photosynthetic prokaryotes."
Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979 (1990).
[2]
FUNCTION.
MEDLINE=92357711; PubMed=1495965;
Math S.K., Hearst J.E., Poulter C.D.;
"The crtB gene in Erwinia herbicola encodes geranylgeranyl
diphosphate synthase".
Proc. Natl. Acad. Sci. U.S.A. 89:6761-6764 (1992).
-!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
phytoene.
-!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
prephytoene diphosphate.
-!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
phytoene.
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; M38423; AAA24821.1; -
EMBL; M87280; AAA64582.1; -
PIR; E39273; E39273.
InterPro; IPR002060; Squ/phyt_synthes.

InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00434; SOS_PSY; 1.
DR PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_2; 1.
DR PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_1; 1.
KW Multifunctional enzyme; Carotenoid biosynthesis; Transferase.
SQ SEQUENCE 309 AA; 34123 MW; B64235C9E0C06F38 CRC64;

Query Match 67.0%; Score 1028; DB 1; Length 309;
Best Local Similarity 64.9%; Pred. No. 1.4e-94;
Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

1 MAVGSKSPATASTLFAKTRRSVLMYAWCRHCDVDVDDQTLGFGFADQSSQWPERLQO 60
14 MAVGSKSPATASTLFAKTRRSVLMYAWCRHCDVDVDDQTLGFGFADQSSQWPERLQO 73
61 LEMKTRQAYAGSQHHEPAPAFQEVAMAHDIAPAYAFDHLGFGFADQSSQWPERLQO 120
74 LEMKTRQAYAGSQHHEPAPAFQEVAMAHDIAPAYAFDHLGFGFADQSSQWPERLQO 133
121 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQCRCYLPESWL 180
134 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQCRCYLPESWL 193
181 EESGLTKANYAAPENRQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 240
194 ODAGLTPTAARENRAALRAVRAELIDRAEPIYSSQAGLHDLPPRCWATATARSVYR 253
241 KIGVKVEQAGQAWDRQSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
254 EIGIKVKAAGSADWRQSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 309

RESULT 3
CRTB_AGRAU STANDARD; PRT; 301 AA.
ID CRTB_AGRAU
AC P54975;
DT 01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene synthase (EC 2.5.1.-).
GN CRTB.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=44153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96062243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiura S., Saito T.,
Ohtani T., Miki W.;
"Structure and functional analysis of a marine bacterial carotenoid
biosynthesis gene cluster and astaxanthin biosynthetic pathway
proposed at the gene level".
J. Bacteriol. 177:6575-6584 (1995).
-!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate
to phytoene.
-!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
prephytoene diphosphate.
-!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
phytoene.
-!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
biosynthetic pathway.
-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; D58420; BAA09595.1; -
InterPro; IPR002060; Squ/phyt_synthes.


```

C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
C Rhodobacteraceae; Rhodobacter.
X NCBI_TaxID=1061;
N SEQUENCE FROM N.A.
X STRAIN=SE11003 / St. Louis, and BEC404;
MEDLINE=89313663; PubMed=2747617;
A Armstrong G.A., Alberti M., Leach P., Hearst J.E.;
I "Nucleotide sequence, organization, and nature of the protein
I products of the carotenoid biosynthesis gene cluster of Rhodobacter
I capsulatus.";
L Mol. Gen. Genet. 216:254-268 (1989).
C -1- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
C phytoene.
C -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
C prephytoene diphosphate.
C -1- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
C phytoene.
C -1- PATHWAY: Carotenoid and chlorophyll biosynthesis.
C -1- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; X52291; CAA36534.1; -.
C PIR; S04403; S04403.
C InterPro; IPR002060; Sm/phyt synthase.
C InterPro; IPR008949; Terpenoid_synth.
C Pfam; PF00494; SQS_PSI; 1.
C PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
C PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
C Multifunctional enzyme; Carotenoid biosynthesis; Transferase;
C Chlorophyll biosynthesis.
C SEQUENCE 339 AA; 37301 MW; 6D7052994934BF4F CRC64;
Query Match 25.6%; Score 393.5; DB 1; Length 339;
Best Local Similarity 35.6%; Pred. No. 9.1e-28;
Matches 103; Conservative 44; Mismatches 125; Indels 17; Gaps 6;
Y 4 GSKSFATATLFDKTRRSVLMYAWCKHCDVDVDDQTLGPHADQPSQMPQRLQOLEM 63
Y 17 GSYSFHANSRVLPARVDPALALYAFCRVADDEVE--VGAPRDKAAVL-----KLGD 68
Y 64 KTRQAYAGSQMHPEP---AFPA--FOEVMAHDIAPAYAFPHLGFADVDVRETRYLIDDTL 119
Y 69 RLEDIYAGRPRNAPSDEAFAAVVEEFEMPRELPEAL----LEGFAWDABGRWTHLSVQV 124
Y 120 RYCVHVAGVGLMMAQIMGVYRDNATLDRACDGLAFOLTNIAARDIVDDAQVGRCYLPESW 179
Y 125 AYSARVAAAAGAMCYLMVRVNPDALARACDGLAQMSNIARDVGEDARGLRFLPTDM 184
Y 180 LEEBGL-TKANYAAPENQALSRVAGRLVREABPYVSSMAGLAQLPLRSAAWATAKOV 238
Y 185 MVEEGIDPQAFADPQPTKGIRRVTERLNRADRLYRAATGVRLLPFFDCRPGIMAAGKI 244
Y 239 YRKIGVKEQAGKQAWDHROSTSTAETKLTLLLTASGOAVTSRMKTYPPR 287
Y 245 YAAIGAELVAKAYKNITRRAHTTKGRKLWLVANSAMSTATSMPLSPR 293
RESULT 6
PSY_NARPS STANDARD; PRT; 423 AA.
AC PS3797;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE
GN GN
OS Narcissus pseudonarcissus (Daffodil).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
OC Narcissus.
OX NCBI_TaxID=39639;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Paracolla;
RA Schledz M., Beyer P.;
RT "Nucleotide sequence of a Narcissus pseudonarcissus cDNA for phytoene
RT synthase.";
RL (in) Plant Gene Register PGR95-122.
CC -1- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
CC phytoene.
CC -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
CC prephytoene diphosphate.
CC -1- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
CC phytoene.
CC -1- PATHWAY: Carotenoid biosynthesis.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78814; CAA55391.1; -.
CC PIR; S54135; S54135.
CC InterPro; IPR002060; Sm/phyt synthase.
CC InterPro; IPR008949; Terpenoid_synth.
CC Pfam; PF00494; SQS_PSI; 1.
CC PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
CC PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
CC Multifunctional enzyme; Carotenoid biosynthesis;
CC Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.
CC TRANSIT 137 423 CHLOROPLAST (POTENTIAL).
CC CHAIN 137 423 PHYTOENE SYNTHASE.
CC SEQUENCE 423 AA; 47805 MW; 8E76BA35312B73DF CRC64;
Query Match 21.4%; Score 328.5; DB 1; Length 423;
Best Local Similarity 27.9%; Pred. No. 7.8e-22;
Matches 81; Conservative 58; Mismatches 132; Indels 19; Gaps 5;
QY 5 SKSFATATLFDKTRRSVLMYAWCKHCDVDVDDQTLGPHADQPSQMPQRLQOLEM 64
QY 140 AKPYLGTLLMTERRRAIWAIVWCERTDELVDGHN-----ASHITPSALDRWEAR 191
QY 65 TRQAYAGSQMHPEPAPAFQEVMAH--DIAPAYAFPHLGFADVDVRETRYLIDDTLRYC 122
QY 192 LEDLFAG-RPYDNFDALSDTVSRFPYDIQF--FMDVGEGRMDLKKSRKYNFDELYLYC 248
QY 123 YHVAGVGLMMAQIMGV-----RDNATLDRACDGLAFOLTNIAARDIVDDAQVGRCYLP 176
QY 249 YVAGTVGLSVVPWGLPAPESLAEASVYNAALAGIANQLTNILDEDVGEDARRGRIYLP 308
QY 177 ESWLEEBGLTKANYAAPENQALSRVAGRLVREABPYVSSMAGLAQLPLRSAAWATAK 236
QY 309 QDELAELAGLSDEDVFTGKVTDKWRSFMKRIKARTFFEQAEKGVTELSQASRPWWASL 368
QY 237 QVYRKIGVKEQAGKQAWDHROSTSTAETKLTLLLTASGOAVTSRMKTYPP 286
QY 369 LLYRQILDEIANDYNNFTKRAVYVSKVKRLAALPLAYTKSLILPLSLRPP 418
RESULT 7
PSY_CAPAN

```

PSY CAPAN STANDARD; PRT; 419 AA.
P37272;
01-OCT-1994 (Rel. 30, Created)
01-OCT-1994 (Rel. 30, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene synthase, chloroplast precursor (BC 2.5.1.-).
PSY1.
Capsicum annuum (Bell pepper).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
lamiales; Solanales; Solanaceae; Capsicum.
NCBI_TaxID=4072;
[1]
SEQUENCES FROM N.A.
STRAIN=CV. Lamuyo;
MEDLINE=94071905; PubMed=8250898;
Roemer S., Huquaney P., Bouvier Z., Camara B., Kuntz M.;
"Expression of the genes encoding the early carotenoid biosynthetic
enzymes in Capsicum annum.";
Biochem. Biophys. Res. Commun. 196:1414-1421 (1993).
-!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
phytoene.
-!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
prephytoene diphosphate.
-!- SUBUNIT: Monomer.
-!- SUBCELLULAR LOCATION: Chloroplast.
-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X68017; CAA48155.1; -
InterPro; IPR002060; Squ/phyt synthase.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00494; SOS_PSY; 1.
PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
Multifunctional enzyme; Carotenoid biosynthesis;
Isoprene biosynthesis; Transferrase; Chloroplast; Transit peptide.
TRANSIT 1 141 CHLOROPLAST (POTENTIAL).
CHAIN 142 419 PHYTOENE SYNTHASE.
SEQUENCE 419 AA; 47126 MW; 9167953D6EA92982 CRC64;
Query Match 20.2%; Score 309.5; DB 1; Length 419;
Best Local Similarity 28.2%; Pred. No. 3.8e-20;
Matches 81; Conservative 56; Mismatches 131; Indels 19; Gaps 6;

5 SKSFATATLFDKTRRSVLMYAWCHRCDDVIDDQTLGFHADQPSQMPQRLQOLEMK 64
144 AKTFVLGTLMTTPERRKAIWAIYWCRTDELVDGPN-----ASHITPAALDRWE 195
65 TRQAYAGSQMHPEAPAPQEVAMAH--DIAPAYAFDHLGFMADVRETRVTLTDDTLRYC 122
196 LEDVPSG-RPFDMLDAALS DTVSKFPVDIQPFR--DMIEGRMDLRKSRVNFDELYLC 252
123 YHVAGVGLXMAQIMGV--RDNATLD---RACDLGLAFQLTNIARDIVDDAOGVRCYLP 176
253 YVAVGTGLASVPIGTAPESSKATTESVYNALALGIANQITNLRVDGDAARGRVYLP 312
177 ESWLEEGTLKNTAAPPNRQALSGIAGRLVREAPYVSSMAGLAQLPLRSAAWATAK 236
313 QDELAQAAGLSDIEDIFAGRVTDKWRIFMKKQIQARKEFDRAEKGVTELSAASRPVLASL 372
237 QVYRKIGVKGVQAGKQANDHQSTSTARKTLLILLITASGQVTSRMT 283

Db 373 LLYRRILDBEANDYNNFTKRAYVSKPKKLLIALPIAYAKSLVPSTRT 419
RESULT 8
PSY2_LYCES
ID PSY2_LYCES STANDARD; PRT; 310 AA.
AC P37273;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene synthase 2, chloroplast precursor (BC 2.5.1.-) (Fragment).
GN PSY2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=94064645; PubMed=8245008;
RA Bartley G.E., Scolnik P.A.;
RT "cDNA cloning, expression during development, and genome mapping of
PSY2, a second tomato gene encoding phytoene synthase.";
RL J. Biol. Chem. 268:25718-25721 (1993).
CC -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
phytoene.
CC -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
prephytoene diphosphate.
CC -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
phytoene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- DEVELOPMENTAL STAGES: In mature leaves.
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L23424; AAA34187.1; -
PIR; A49558; A49558.
InterPro; IPR002060; Squ/phyt synthase.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00494; SOS_PSY; 1.
PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
Multifunctional enzyme; Carotenoid biosynthesis; Multigene family;
Isoprene biosynthesis; Transferrase; Chloroplast; Transit peptide.
NON_TER 1 1
TRANSIT 26 310 CHLOROPLAST (POTENTIAL).
FT CHAIN 26 310 PHYTOENE SYNTHASE 2.
SEQUENCE 310 AA; 35224 MW; 8EAF60FELD9F4CF5 CRC64;
Query Match 20.0%; Score 307.5; DB 1; Length 310;
Best Local Similarity 28.9%; Pred. No. 4e-20;
Matches 84; Conservative 53; Mismatches 129; Indels 25; Gaps 7;

5 SKSFATATLFDKTRRSVLMYAWCHRCDDVIDDQTLGFHADQPSQMPQRLQOLEMK 64
28 AKTFVLGTLMTTPERRKAIWAIYWCRTDELVDGPN-----ASHITPAALDRWE 79
65 TRQAYAGSQMHPEAPAPQEVAMAH--DIAPAYAFDHLGFMADVRETRVTLTDDTLRYC 122
80 LEDIFNG-RPFDMLDAALS DTVSKFPVDIQPFR--DMVEGRMDLRKSRVNFDELYLC 136
123 YHVAGVGLXMAQIMGV--RDNATLD---RACDLGLAFQLTNIARDIVDDAOGVRCYLP 176

b 137 YVAGTVGLMSVPMGIAPEKATTSVTVNLAALGIANQLNLRDVGEDARRGRVYLP 196
Y 177 ESMLEEGTKANYAAPENRQALSRIAGSLVREAEPPYVSSMAGLAQLPLRSAMAIATAK 236
b 197 ODELAQAGLSDEIPAGKVTDKWIFPKWQIQKARFFDEAEKGVTLSASASWPVLASL 256
Y 237 QVTRKIGVKEQAGKQANDHROSTSTAEKLTLLLTASGOAVTSRKMTYPPR 287
b 257 LLYRKILDEIANDYNNFTTRAYVSPKXLLTLPIAYASLSV-----PPK 301

RESULT 9

SY DAUCA
D - PSI DAUCA STANDARD; PRT; 398 AA.
C QSSUS;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene synthase, chloroplast precursor (EC 2.5.1.-).
N PSI.
S Daucus carota (Carrot).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C campanulids; Apiales; Apiaceae; Apiaceae; Scandiceae; Daucinae;
C Daucus.
X NCBI_TaxID=4039;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=ROOT;
A Uno T., Sankawa U.;
T "Daucus carota phytoene synthase.";
L Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
L - FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
C phytoene.
C - CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
C prephytoene diphosphate.
C - CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
C phytoene.
C - PATHWAY: Carotenoid biosynthesis.
C - SUBUNIT: Monomer (By similarity).
C - SUBCELLULAR LOCATION: Chloroplast.
C - SIMILARITY: Belongs to the phytoene/squalene synthetase family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C
C EMBL; AB032797; BAA84763.1; -
C InterPro; IPR002060; Squ/phyt synthase.
C InterPro; IPR008949; Terpenoid synth.
C Pfam; PF00494; SQS_PSY_1
C PROSITE; PS01044; SQUALEN_PHTOEN_SYN_1; 1.
C PROSITE; PS01045; SQUALEN_PHTOEN_SYN_2; 1.
C Multifunctional enzyme; Carotenoid biosynthesis;
C Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.
C TRANSIT ? 398 PHYTOENE SYNTHASE.
C CHAIN ? 398
C SEQUENCE 398 AA; 45199 MW; 1D1E043824730615 CRC64;

Query Match 20.0%; Score 307.5; DB 1; Length 398;
Best Local Similarity 25.5%; Pred. No. 5.4e-20;
Matches 85; Conservative 57; Mismatches 113; Indels 33; Gaps 9;

Y 5 SKSFATATLTDKTRSLVLMYAWCRCHCDVDDQTLGFHADQPSQSQPEQLQLEMK 64
b 122 AKTYLGLTLPERRAVAVIYWCRTSLVDGPN-----ASHITPKALDRWEKR 173
Y 65 TRQYAGSQMHEPAFAQFQVAMAH--DIPAPAFDHEGFAMDVRETRYLTLDOTLRVC 122

Db 174 LNDLFDG-QPYDMYDAALADTYSTYVPDIQPFK--DMIDGEMDLKKSRYQTFDBSLYLC 230
Qy 123 YFVAGVGLMMAQIMGV--RDNATLD-----RACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
Db 231 YVAGTVGLMSVPMGIAPEKATTSVTVNLAALGIANQLNLRDVGEDARRGRVYLP 290
Qy 177 ESMLEEGTKANYAAPE-----NRQALSRIAGSLVREAEPPYVSSMAGLAQLPLRSA 229
Db 291 QBELKLAGIT-----PEVIFKGVTDKRSFPMKQO-IKRARMFFDEAEKGVAELSSASR 343
Qy 230 WAIATATQVYRKIGVKEQAGKQANDHROSTSTAEKLTLLLTASGOAV 277
Db 344 MPVWASLLYKQILDIAEANDYNNFTTRAYVSPKXLLTLPIAYASLSRAL 391

RESULT 10

PSY ARATH
ID - PSI ARATH STANDARD; PRT; 422 AA.
AC P37271; O22375; Q8LE86;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phytoene synthase, chloroplast precursor (EC 2.5.1.-).
GN PSI OR PSY OR ATSGI7230 OR MKP11.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=94286755; PubMed=8016277;
RA Scolnik P.A., Bartley G.B.;
RT "Nucleotide sequence of an Arabidopsis cDNA for phytoene synthase.";
RL plant Physiol. 104:1471-1472(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX Castagnano P., Giuliano G.;
RT "Sequence of the phytoene synthase gene of Arabidopsis.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT P1 clones";
RL DNA Res. 4:215-230(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.A.;
RT "Full-length cDNA from Arabidopsis thaliana";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kaniya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,

Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.:
"Empirical analysis of transcriptional activity in the Arabidopsis
genome.";
Science 302:842-846(2003).

-!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
phytoene.
-!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
prephytoene diphosphate.
-!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
phytoene.
-!- PATHWAY: Carotenoid biosynthesis.
-!- SUBUNIT: Monomer (By similarity).
-!- SUBCELLULAR LOCATION: Chloroplast.
-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

EMBL; L25812; AAA32836.1; .
EMBL; AF009954; AAB65697.1; .
EMBL; AB005238; BAB10510.1; .
EMBL; AY085565; AAM62787.1; .
EMBL; BT000450; AAN17427.1; .
EMBL; BT002084; AAN72095.1; .
InterPro; IPR002060; Squ/phyt_synthase.
InterPro; IPR008949; Terpenoid_synth.
Pfam; PF00494; SQS_PSY; 1.
PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
Multifunctional enzyme; Carotenoid biosynthesis;
Isoprene biosynthesis; Transferrase; Chloroplast; Transit peptide.
CHAIN 71 422 CHLOROPLAST (POTENTIAL).
CONFLICT 60 60 R -> M (IN REF. 4).
CONFLICT 128 128 L -> LV (IN REF. 1).
CONFLICT 143 143 A -> P (IN REF. 1).
SEQUENCE 422 AA; 47486 MW; C44FOA512F2DD31E CRC64;

Query Match 20.0%; Score 307; DB 1; Length 422;
Best Local Similarity 27.8%; Pred. No. 6.5e-20;
Matches 81; Conservative 58; Mismatches 130; Indels 22; Gaps 7;

5 SKSFATSTLFDKTRRSVLMYAKRCHDDVIDDQTLGFHADQPSQMPQRLOQLEMK 64
143 AKTFYLGTLMTPEPKKALWAIYVWCRTDDELVDGPN-----ASHITPMALDRWEAR 194
65 TRQAVAGSQMHEPAPAFQEVAMAH--DIAPAYAFDHLGFMADVRETRYLTLDLRYC 122
195 LEDLFRG-RPFDMLDAALADTVARYVPDIQPPR--DMIEGMRMDLKKSRQNFDDLYLC 251
123 YHVAGVVGGLMAQIMGV--RDNATLD----RACDGLAFQLTNIARDIVDDAQVGRCYLP 176
252 YVYAGTVGLMSVPVWGIDPKSKATTESVYNAALAGIANQLTNILRDVGEDARRGRVLP 311
177 ESWLEBEGITKANYAAPENRQALSRIAGLRVREAPFYVSSWAGLAQLPLRSAMAIATAK 236
312 QBELAQAGLSDEIDFAGKVTQWRNFMKQKRAKRVFFDEAKGVTELSAASRPFWASL 371
237 QVYRKIGVKVQAGKQAWDHROSTSTAEKLTLLTLLTASQAVTSRKVTYPPR 287
372 LLYRQILDBEANDYNNFTKRAYGVKGKLLALPLAYAKSV---LNTSSSR 419

RESULT 11
SY MAIZE
C PSY MAIZE PRT; 410 AA.
C P49085;
I 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene synthase, chloroplast precursor (EC 2.5.1.1-).
GN Y1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8722797;
RX MEDLINE=96304610; PubMed=8722797;
RA Buckner B., Sammguel P., Janick-Buckner D., Bennettzen J.L.;
RT "The Y1 gene of maize codes for phytoene synthase.";
RL Genetics 143:479-488(1996).
CC -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
phytoene.
CC -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
prephytoene diphosphate.
CC -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
phytoene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

EMBL; U32636; AAB60314.1; .
PIR; S68307; S68307.
DR MaizeDB; 66643; .
DR InterPro; IPR002060; Squ/phyt_synthase.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00494; SQS_PSY; 1.
DR PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
DR PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
KW Multifunctional enzyme; Carotenoid biosynthesis;
KW Isoprene biosynthesis; Transferrase; Chloroplast; Transit peptide.
FT TRANSIT ? 410 CHLOROPLAST (POTENTIAL).
FT CHAIN 344 344 N -> T (IN ALLELE B73).
FT VARIANT 344 344 N -> T (IN ALLELE B73).
FT SEQUENCE 410 AA; 46481 MW; 21070A33624ED79 CRC64;

Query Match 19.7%; Score 302.5; DB 1; Length 410;
Best Local Similarity 28.1%; Pred. No. 1.6e-19;
Matches 79; Conservative 53; Mismatches 130; Indels 19; Gaps 6;

5 SKSFATSTLFDKTRRSVLMYAKRCHDDVIDDQTLGFHADQPSQMPQRLOQLEMK 64
130 AKTFYLGTLMTPEPKKALWAIYVWCRTDDELVDGPNY-----ITPTALDRWEKR 181
65 TRQAVAGSQMHEPAPAFQEVAMAH--DIAPAYAFDHLGFMADVRETRYLTLDLRYC 122
182 LEDLFTG-RPYDMLDAALSDTISRFPDIQPPR--DMIEGMRSDLRKTRYNFDELNYC 238
123 YHVAGVVGGLMAQIMGV--RDNATLD----RACDGLAFQLTNIARDIVDDAQVGRCYLP 176
239 YVYAGTVGLMSVPVWGIDPKSKATTESVYNAALAGIANQLTNILRDVGEDARRGRVLP 298
177 ESWLEBEGITKANYAAPENRQALSRIAGLRVREAPFYVSSWAGLAQLPLRSAMAIATAK 236
299 QBELAQAGLSDEIDFAGKVTQWRNFMKQKRAKRVFFDEAKGVTELSAASRPFWASL 358
237 QVYRKIGVKVQAGKQAWDHROSTSTAEKLTLLTLLTASQAV 277
359 LLYRQILDBEANDYNNFTKRAYGVKGKLLALPVYAKSL 399

RESULT 12
SYNOPSIS
D PSYL LUCES STANDARD; PRT; 412 AA.
C P08136;
T 01-AUG-1988 (Rel. 08, Created)
T 01-OCT-1994 (Rel. 30, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytene synthase 1, chloroplast precursor (EC 2.5.1.-) (Fruit ripening specific protein PTOMS).
N PSYL OR PTOMS.
S Lycopersicon esculentum (Tomato).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C Lamiales; Solanales; Solanaceae; Solanum.
X NCBI_TaxID=4081;
N [1]
P SEQUENCE FROM N.A. AND CHARACTERIZATION.
X MEDLINE=92184738; PubMed=1544888;
A Bartley G.E., Viltanen P.V., Bacot K.O., Soolnik P.A.;
T "A tomato gene expressed during fruit ripening encodes an enzyme of
T the carotenoid biosynthesis pathway."
L J. Biol. Chem. 267:5036-5039(1992).
N [2]
P SEQUENCE FROM N.A.
C STRAIN=cv. Marmande;
X MEDLINE=94071305; PubMed=8250898;
A Roemer S., Hugueney P., Bouvier F., Camara B., Kuntz M.;
T "Expression of the genes encoding the early carotenoid biosynthetic
T enzymes in Capsicum annum."
L Biochem. Biophys. Res. Commun. 196:1414-1421(1993).
N [3]
P SEQUENCE FROM N.A.
C STRAIN=cv. Ailsa Craig;
X MEDLINE=88096591; PubMed=3697097;
A Ray J., Bird C.R., Maunders M., Grierson D., Schuch W.;
T "Sequence of pPOMS, a ripening related cDNA from tomato."
L Nucleic Acids Res. 15:10587-10587(1987).
N [4]
P SEQUENCE FROM N.A.
C STRAIN=cv. Ailsa Craig; TISSUE=Leaf;
X MEDLINE=92322971; PubMed=1623189;
A Ray J., Moureau P., Bird C., Bird A., Grierson D., Maunders M.,
A Truesdale M., Bramley P., Schuch W.;
T "Cloning and characterization of a gene involved in phytoene
T synthesis from tomato."
L Plant Mol. Biol. 19:401-404(1992).
N [5]
P FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
C phytoene.
C -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
C prephytoene diphosphate.
C -1- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
C phytoene.
C -1- PATHWAY: Carotenoid biosynthesis.
C -1- SUBUNIT: Monomer (By similarity).
C -1- SUBCELLULAR LOCATION: Chloroplast.
C -1- DEVELOPMENTAL STAGE: In seedlings and in late stages of fruit
C ripening.
C -1- INDUCTION: By fruit ripening.
C -1- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C
C EMBL; M84744; AAA34153.1; -
C EMBL; Y00521; CAA68575.1; -
C EMBL; X60441; CAA42969.1; -
C EMBL; A21360; CAA01548.1; -

DR PIR; A42102; A42102.
DR PIR; S22474; S22474.
DR InterPro; IPR002060; Squ/phyt synthase.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00494; SQS_PSY; 1.
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
KW Multifunctional enzyme; Carotenoid biosynthesis; Multigene family;
KW Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.
FT TRANSIT 1 129 CHLOROPLAST (POTENTIAL).
FT CHAIN 130 412 PHYTOENE SYNTHASE 1.
FT CONFLICT 389 407 KLALPIAYAKSLVPPKT ->
FT CONFLICT 389 405 QVDCITYCKICKSCASY (IN REF. 3).
FT CONFLICT 389 405 KLALPIAYAKSLVPPKT ->
FT CONFLICT 389 405 QVDCITYCKICKSCASY (IN REF. 4).
SQ SEQUENCE 412 AA; 46615 MW; DBED05945D6F4C9A CRC64;

Query Match 19.6%; Score 300.5; DB 1; Length 412;
Best Local Similarity 29.0%; Pred. No. 2.4e-19;
Matches 83; Conservative 51; Mismatches 135; Indels 17; Gaps 5;

QY 5 SKGFATASTLFDKTRRSVLMYAWCRHCDVDVDTLGFHADQFSSQMPQRLOOLEMK 64
DB 132 AKTNLGLTMTPEERRAIWAIYVWCRTDVLVDFN-----ASYITPAALDRWENR 183

QY 65 TRQAYAGSOMHEPAFAFQCEVA-MAHDIAPAYAFDHLGCFAMDVRETRYLTLDLTLYCY 123
DB 184 LEDVNGRPPDMLDGLSDTVSNFPYDIOPFR--DMIBGEMDLKRSYKFNDELYLYCY 241

QY 124 HVAGVGLMMAQIMGV--RONATLD----RACDLGLAFQLTNIARDIVDDAQVGRCYLPE 177
DB 242 YVAGTVGLMSVPIMGIAPEASKATTESVYNAALALGIANQLTNILRDVGEDARGRGVYLPQ 301

QY 178 SWLEEGLTAKYAAAPENRQALSRAGLRVRAEPYVYSSMAGLAQLPLRSAAWATATAKQ 237
DB 302 DELAQGLSDEDFAGRVTDKWRIFPKQIHARXFFDEAEKGVTELSASRFPWVASIV 361

QY 238 VYRKIGVQVQAKQAWDHQSTSTAEKLTLLLTASGQAVTSMKT 283
DB 362 LYRKILDEIEANDYNNFTKRVYVSKSKLIALPIAYAKSLVPPTKT 407

RESULT 13
PSY_NEUCR
ID PSY_NEUCR STANDARD; PRT; 602 AA.
AC P37293;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DR Phytoene synthase (EC 2.5.1.-) (Albino-2 protein).
GN AL-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=94216317; PubMed=8163509;
RA Schmidhauser T.J., Lauter F.-R., Schumacher M., Zhou W., Russo V.E.A.,
RA Yanofsky C.;
RT "Characterization of al-2, the phytoene synthase gene of Neurospora
RT crassa. Cloning, sequence analysis, and photoregulation."
RL J. Biol. Chem. 269:12060-12066(1994).
CC -1- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
CC phytoene.
CC -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
CC prephytoene diphosphate.
CC -1- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
CC phytoene.
CC -1- PATHWAY: Carotenoid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- INDUCTION: By blue light.

```

-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L27652; AA19428.1; -.
PIR; A53583; A53583.
InterPro; IPR002060; Squ/phyt synthetase.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00494; SQS_PSY; 1.
PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
Multifunctional enzyme; Carotenoid biosynthesis; Transmembrane;
Isoprene biosynthesis; Transferase.
TRANSMEM 6 26 POTENTIAL.
TRANSMEM 30 50 POTENTIAL.
TRANSMEM 76 96 POTENTIAL.
TRANSMEM 118 138 POTENTIAL.
TRANSMEM 146 166 POTENTIAL.
TRANSMEM 168 188 POTENTIAL.
TRANSMEM 230 250 POTENTIAL.
SEQUENCE 602 AA; 58927 MW; DDFBFD725007B5C2 CRC64;

Query Match 19.3%; Score 297; DB 1; Length 602;
Best Local Similarity 28.6%; Pred. No. 7.9e-19;
Matches 88; Conservative 57; Mismatches 105; Indels 58; Gaps 10;

Y 5 SKGFATASTLFDKATRSVLMYAMCRHCDVDVDD-----QTLGFHA-----DQ 48
b 295 SRSFYLANSLFSGRLAIDILLYSLFCLLDVDDAKSRREVLSTAKLNHFLDHYKDA 354
Y 49 PSSQMEQRLQLEMTKTRQAYAGSOMHEPAPAFQEVAM--AHDIAPAYAFDHLGFMAD 106
b 355 DATEDPKKAERDAVYKTAFFP-----CAYQALHLLPHLPKPLDYLDLNGFEMD 406
Y 107 VRETRYLTLD-----DTRYCYHVAGVVG-LMMAQI-----MGVRDNATLD 146
b 407 SQTFPHGTSIDLOVPIADDDXDLNLYAVAGTVGELCIAIIVHCLPDMSDTKRLE 466
Y 147 -RACDGLAFLNTIARDIVDDAOGRCYLPRESLEEGLT-KANYAAPENKQALSRIAG 204
b 467 TAACRMGIALQVNIARDIVDARIGRVLPPTWLKXGLTHKWLENBEGPEVIERMR 526
Y 205 RLVRAEAPYVSSMAGLAQLPLRSAAWAIATAKOVYRKIG-----VKVBOAGKQAW 254
b 527 RLLENAPELYGGARPQWQIPSEARGPMIGAVENTYMAIGRVLRERKEGTVFVYRMEGRATV 586
Y 255 DHRQSTST 262
b 587 PKRRRLST 594

RESULT 14
SY_CUCURB STANDARD; PRT; 422 AA.
D P49293;
T 01-FEB-1996 (Rel. 33, Created)
T 01-FEB-1996 (Rel. 33, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene synthase, chloroplast precursor (EC 2.5.1.-) (MEL5).
N PSY.
S Cucumis melo (Muskmelon).
C Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
C Cucurbitales; Cucurbitaceae; Cucumis.
X NCBI_TaxID=3656;
N [1]
P SEQUENCE FROM N.A.

```

```

RC STRAIN=cv. Cantaloup Charentais; TISSUE=Pericarp;
RX MEDLINE=95284364; PubMed=7766896;
RA Karyouni Z., John I., Taylor J.E., Watson C.F., Turner A.J.,
RA Grierson D.;
RT isolation and characterisation of a melon cDNA clone encoding
RT phytoene synthase";
RL Plant Mol. Biol. 27:1153-1162(1995).
CC -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
CC phytoene.
CC -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
CC prephytoene diphosphate.
CC -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
CC phytoene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z37543; CA85775.1; -.
DR PIR; S56668; S56668.
DR InterPro; IPR002060; Squ/phyt synthetase.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00494; SQS_PSY; 1.
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
KW Multifunctional enzyme; Carotenoid biosynthesis;
KW Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.
PT TRANSIT 1 83 CHLOROPLAST (POTENTIAL).
PT CHAIN 84 422 PHYTOENE SYNTHASE.
SQ SEQUENCE 422 AA; 47392 MW; 259124B3AD4642B0 CRC64;

Query Match 18.6%; Score 286; DB 1; Length 422;
Best Local Similarity 26.5%; Pred. No. 4.9e-18;
Matches 76; Conservative 56; Mismatches 125; Indels 30; Gaps 6;

QY 5 SKGFATASTLFDKATRSVLMYAMCRHCDVDVDDQTLGFHADQSSQMEQRLQLEMK 64
Db 148 AKTYLGTMLNTPERQKAIWAIYVWCRDTELDVDPN-----ASHITPTALDRWEAR 199
QY 65 TRQAYAGSQMHEPAPAFQEVAMAH-----DIAPAYAFDHLGFMADVRETRYLTLD 117
Db 200 LEEFQORPD-----MLDALADTVTKFPVDIQPK--DMIEGWMDLKRSYKNPDE 251
QY 118 TLRYCYHVAGVVGVMVAQINGV-----RDNATLDRACDLGLAFL-TNIAARDIVDDAQV 170
Db 252 LDYCYVYVAGTVGLNSVPVNGIAPESQASTESVYNAALALGIANQAPPNILRDVGEDARR 311
QY 171 GRCYLPESMLEEGLTKANYAAPENKQALSRIAGRLVREAEFYVSSMAGLAQLPLRSAAW 230
Db 312 GRITLPDELAQAGLSDEDIFAGRVTDKWKNFMKNQIKKARMFFDEAEKGVLELNKASRW 371
QY 231 AIATAKQVYRKIGVKVEQAGKQAWDHRQSTSTAEKLTLLLTASQAV 277
Db 372 PVWASULLYRQILDEIANDYDNFTKRAYVSKAKKILALPMAYGRAL 418

RESULT 15
ID_CRTY SPIPL STANDARD; PRT; 309 AA.
AC O07333;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene synthase (EC 2.5.1.-).
GN CRTB OR PSY.

```

```

XS Spirulina platensis.
X  Bacteria; Cyanobacteria; Oscillatoriaceae; Spirulina.
X  NCB1_TaxID=1156;
X  [1]
X  SEQUENCE FROM N.A.
X  STRAIN=IAM M-135;
X  KAWATA Y., YANO S., KOJIMA H.;
X  Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
X  L  -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
X  C  phytoene.
X  C  -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
X  C  prephytoene diphosphate.
X  C  -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
X  C  phytoene.
X  C  -!- PATHWAY: Carotenoid biosynthesis.
X  C  -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
X  C  -----
X  C  This SWISS-PROT entry is copyright. It is produced through a collaboration
X  C  between the Swiss Institute of Bioinformatics and the EMBL outstation -
X  C  the European Bioinformatics Institute. There are no restrictions on its
X  C  use by non-profit institutions as long as its content is in no way
X  C  modified and this statement is not removed. Usage by and for commercial
X  C  entities requires a license agreement (See http://www.isb-sib.ch/announce/
X  C  or send an email to license@isb-sib.ch).
X  C  -----
X  C  EMBL; AB001284; BAA20384.1; -
X  C  InterPro; IPR002060; Squ/phyt_synthase.
X  C  InterPro; IPR008949; Terpenoid_synth.
X  C  Pfam; PF00494; SQS_PSY; 1.
X  C  PROSITE; PS01044; SQUALEN_PHYTOEN SYN 1; 1.
X  C  PROSITE; PS01045; SQUALEN_PHYTOEN SYN 2; 1.
X  C  Multifunctional enzyme; Carotenoid biosynthesis; Transferase.
X  C  SEQUENCE 305 AA; 35345 MW; C42A1C6431604C75 CRC64;
X  C  -----
Query Match      18.2%; Score 280; DB 1; Length 309;
Best Local Similarity 23.7%; Pred. No. 1.1e-17;
Matches 86; Conservative 44; Mismatches 128; Indels 32; Gaps 8;
2Y 5 SKSFATASTLPDAKTRSVLMYAWCHDCDDVIDDQTGLGHADQPSSOMPEORLQOLEMK 64
2D 31 AKTYLGLQMLAKRQAIWAIYVWCRTBELVDGFM-----ASSTLTLELDHWEHQ 82
2Y 65 TRQVAGSQMHEPAAFOEV--AMAHDIAPAYAFDHLGFMDEVRETRYLTLDLDTLYC 122
2D 83 LESIFAGHPI-EPVDVALVDTLGRFPDLIQPER--DMIAGQMDLSRNYNTFDELNLYC 139
2Y 123 YHVGAVGLMAQIMGV-----RDNATL---DRACDLGLAQLTNIARDIVDDAQ 169
2D 140 YRVAGTVGLMSLAYNGTAEPDLSVPWRDQSIYYPKEEALGLANQLNLRDVGEDAR 199
2Y 170 VRCVLPESWLEEEGLTKA---NYAAPENQALSRIAGRLVREAEPPYVSSMAGLAQLPL 226
2D 200 RGRVLPDLDDLALFNFTYEDLLNGKVDWRWELMRFO---IQARKEYTLAEEGIAALHP 256
2Y 227 RSANAIAATKQVYKIGVKVQKQAWDHQSTSTAEKLTLLITASQQA 276
2D 257 DIRMEVWTALMYRQILDEIERNEYDVFNQRAYVPTWKMMCLPLAQLRA 306

```

Search completed: February 29, 2004, 14:45:14
Job time : 6.98876 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:34:14 ; Search time 8.23902 Seconds
(without alignments)
3455.835 Million cell updates/sec

tle: US-09-941-947A-34

rfect score: 1535
quence: 1 MAVGSKSFATSTLFDKTR.....VTSRMKTPRPAHLWQRP 296

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tal number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1371	89.3	296	2 S52587	prephytoene pyroph
2	1368	89.1	296	2 E37802	crbB protein - Erw
3	1028	67.0	309	2 B39273	geranylgeranyl-dip
4	395.5	25.8	344	2 T30895	prephytoene pyroph
5	394	25.7	355	2 S49621	phytoene synthase
6	393.5	25.6	339	2 S04403	geranylgeranyl-dip
7	385	25.1	355	2 T50746	phytoene synthase
8	362	23.6	325	2 D75466	phytoene synthase
9	328.5	21.4	423	2 S54135	phytoene synthase
10	307.5	20.0	310	2 A49558	phytoene synthase
11	302.5	19.7	410	2 S68307	phytoene synthase
12	300.5	19.6	412	2 A42102	phytoene synthase
13	300	19.5	336	2 S32170	geranylgeranyl-dip
14	297	19.3	602	2 A53583	phytoene synthase
15	286	18.6	412	2 S22474	geranylgeranyl-dip
16	286	18.6	422	2 S56668	phytoene synthase
17	276.5	18.0	425	2 T10702	phytoene synthase
18	266	17.3	310	2 A2035	phytoene synthase
19	265.5	17.3	337	2 S45360	phytoene synthase
20	261.5	17.0	290	2 A81118	phytoene synthase
21	258	16.8	307	2 S20383	probable poly-isop
22	255.5	16.6	290	2 H81902	probable phytoene
23	248.5	16.0	331	2 T36969	probable phytoene
24	228.5	14.9	312	2 T35400	phytoene synthase
25	224.5	14.6	319	2 T46594	probable phyA prot
26	217.5	14.2	302	2 B70735	phytoene synthase
27	214.5	14.0	322	2 H84299	phytoene synthase
28	210.5	13.7	299	2 F69108	phytoene synthase
29	209.5	13.6	322	2 E84320	phytoene synthase

30	201.5	13.1	277	2 G90469	phytoene synthase
31	196.5	12.8	275	2 H81074	probable transfe
32	193.5	12.6	276	2 D81868	squalene synthase
33	171.5	11.2	417	2 F52090	farnesyl-diphospha
34	171.5	11.2	417	2 A45998	farnesyl-diphospha
35	171.5	11.2	417	2 T38245	farnesyl-diphospha
36	165	10.7	362	2 T45141	squalene desaturas
37	164.5	10.7	287	2 G90061	farnesyl-diphospha
38	163.5	10.7	416	2 S52075	farnesyl-diphospha
39	163	10.6	390	2 A84226	farnesyl-diphospha
40	159.5	10.4	416	2 A45105	probable phytoene
41	144.5	9.4	303	2 T35399	phytoene synthase
42	140.5	9.2	274	2 G69837	phytoene synthase
43	135	8.8	299	2 T51118	crbM protein - Sta
44	134	8.7	254	2 A55548	farnesyl-diphospha
45	132.5	8.6	460	2 B48057	

ALIGNMENTS

RESULT 1

S52587

prephytoene pyrophosphate synthase - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 02-Mar-2001

C:Accession: S52587

R:Lin. Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52587

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-296 <LIN>

A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21264.1; PID:G148398

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992

C:Superfamily: Mycobacterium marinum phytoene synthase

Query Match 89.3%; Score 1371; DB 2; Length 296;

Best Local Similarity 88.5%; Pred. No. 7e-116;

Matches 262; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY	1	MAVGSKSFATSTLFDKTRSVLMYVACRCHDDVDDQTLGPHADQPSQPMQEQRLQQ	60
DB	1	MAVGSKSFATSKLFDKTRSVLMYVACRCHDDVDDQTLGPHADQPSQPMQEQRLQQ	60
QY	61	LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR	120
DB	61	LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR	120
QY	121	YCHVAGVVGVLMAQINGVRDNATLDACDGLAFQLTNIARDIVDDAOGVCYLPBSWL	180
DB	121	YCHVAGVVGVLMAQINGVRDNATLDACDGLAFQLTNIARDIVDDAOGVCYLPBSWL	180
QY	181	EEEGTKAYAAPENKQALSRIAGRLVREAPYVSSNAGLAQLPLRSAMAIATAKQYR	240
DB	181	ENEGINKENYAAPENKQALSRIARELVCEAPYLYCATAGLAGLPLRSAMAIATAKQYR	240
QY	241	KIGVKVEQAGKQAWDHROSTSTAETKLTLLLTASQAVTSRMKTYPPRPAHLWQRP	296
DB	241	KIGVKVEQAGTAEWDHROSTTTPKLSLLMASQAVTSRMKTYPPRPAHLWQRP	296

RESULT 2

E37802

crbB protein - Erwinia uredovora

C:Species: Erwinia uredovora

C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 02-Mar-2001

C:Accession: E37802

R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim

J. Bacteriol. 172, 6704-6712, 1990

A:Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio

;Reference number: A37802; MUID:91072214; PMID:2254247

;Accession: B37802

;Status: preliminary

;Molecule type: DNA

;Residues: 1-296 <NIS>

;Cross-references: GB:D90087; NID:g216681; PIDN:BAAI4128.1; PID:d1014831; PID:g216686

;Superfamily: Mycobacterium marinum phytoene synthase

Query Match 89.1%; Score 1368; DB 2; Length 296;

Best Local Similarity 88.9%; Pred. No. 1.3e-115;

Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

Y 1 MAVGSKSFATATLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQYPERLQQ 60

b 1 MAVGSKSFATATLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQYPERLQQ 60

Y 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR 120

b 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR 120

Y 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

b 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

Y 181 EEEGLTKANYAAPENRQALSRVAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 240

b 181 EEEGLTKANYAAPENRQALSRVAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 240

Y 241 KIGVKVQAGKQANDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

b 241 KIGVKVQAGKQANDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

RESULT 3

339273

;Alternate names: geranyl-diphosphate geranylgeranyltransferase (GC 2.5.1.32) - Erwinia herbicola

;Species: Erwinia herbicola

;Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 02-Mar-2001

;Accession: B39273; B33120

;Armstrong, G.A.; Alberti, M.; Hearst, J.E.

;Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990

;Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nor

;Reference number: A39273; MUID:91088634; PMID:2263648

;Accession: B39273

;Molecule type: DNA

;Residues: 1-309 <ARM>

;Cross-references: GB:M38423; NID:g148401; PIDN:AAA24821.1; PID:g148403

;Gene: CrtB

;Superfamily: Mycobacterium marinum phytoene synthase

;Keywords: transferase

Query Match 67.0%; Score 1028; DB 2; Length 309;

Best Local Similarity 64.9%; Pred. No. 6.2e-85;

Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

Y 1 MAVGSKSFATATLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQYPERLQQ 60

b 14 MAVGSKSFATATLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQYPERLQQ 60

Y 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR 120

b 74 LRTLTAAEFGAEMQDPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR 133

Y 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

b 134 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 193

Y 181 EEEGLTKANYAAPENRQALSRVAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 240

b 194 QADGLTPENYAARENRAALVAERLIIDAAEFYIISQAGLHDLPPRCAMAIATARSVYR 253

QY 241 KIGVKVQAGKQANDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

DB 254 EIGIKVRAAGSANDRQHTSKGEKIAMLWAGQVIRAKTTTVPTRPAGLWQRPV 309

RESULT 4

T50895

;prephytoene pyrophosphate synthase [imported] - Rubrivivax gelatinosus

;Species: Rubrivivax gelatinosus

;Date: 21-Jul-2000 #sequence revision 21-Jul-2000 #text_change 02-Mar-2001

;Accession: T50895

;Nagaohima, K.V.; Igarashi, N.; Harada, J.; Nagaehima, S.; Matsura, K.; Shimada, K.

;submitted to the EMBL Data Library, November 1999

;A:Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosyn

;A:Reference number: Z25270

;A:Accession: T50895

;A:Status: preliminary; translated from GB/EMBL/DDBJ

;A:Molecule type: DNA

;A:Residues: 1-344 <NAG>

;A:Cross-references: EMBL:AB034704; PIDN:BA094048.1

;A:Experimental source: strain IL144

;C:Genetics:

;A:Gene: crtB

;C:Superfamily: Mycobacterium marinum phytoene synthase

Query Match 25.8%; Score 395.5; DB 2; Length 344;

Best Local Similarity 36.2%; Pred. No. 7.3e-28;

Matches 106; Conservative 32; Mismatches 140; Indels 15; Gaps 3;

QY 1 MAVGSKSFATATLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQYPERLQQ 60

DB 18 MRGSKSFFAASLLLPQVTRTPATLYAFRVADDAVD-----LSGDPHAAWDE 66

QY 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDOTLR 120

DB 67 LRTLDVAYAGTAPTAADRALASTVHRVYVPRVLLDALLLEGFLWDAGRRYDTIADVEA 126

QY 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

DB 127 YGARVAGTGAAGALMGVRSPOALARACELGVAMQFTNIARDVGDGARDNGRLYLPRWL 186

QY 181 EEEGLTKANYAAPENRQALSRVAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVY 239

DB 187 VEAGLDVDMQLQNPVHCEVAQVTRVRLRAADBELYSEBHGIAALPDCRPAIRARLVY 246

QY 240 KIGVKVQAGKQANDHROSTSTAEKLTLLLTASGOAVTSRMKTY---PPRPA 289

DB 247 AEIGRMLERGDLSVNRVVPVPPARRKAALMARASAAFTNTPGRATISMPPPLPA 299

RESULT 5

S49621

;phytoene synthase - Rhodobacter sphaeroides

;Species: Rhodobacter sphaeroides

;Date: 05-Mar-1995 #sequence revision 12-May-1995 #text_change 02-Mar-2001

;Accession: S49621

;Rieng, H.P.; Cogdell, R.J.; Takaichi, S.; Hunter, C.N.

;submitted to the EMBL Data Library, November 1994

;A:Description: The complete DNA sequence, specific TMS insertion map and gene assignmen

;A:Reference number: S49619

;A:Accession: S49621

;A:Status: preliminary

;A:Molecule type: DNA

;A:Residues: 1-355 <LAN>

;A:Cross-references: EMBL:X82456; NID:g575405; PID:g575408

;C:Genetics:

;A:Gene: crtB

;C:Superfamily: Mycobacterium marinum phytoene synthase

Query Match 25.7%; Score 394; DB 2; Length 355;

Best Local Similarity 36.8%; Pred. No. 1e-27;

Matches 105; Conservative 32; Mismatches 128; Indels 22; Gaps 3;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:51:24 ; Search time 19.6527 Seconds
(without alignments)
3180.293 Million cell updates/sec

file: US-09-941-947A-34

fect score: 1535
quence: 1 MAVGSKSFATASTLFDKTR.....VTSRMKTPRPAHLWQRP 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 211153259 residues

tal number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubaa/US09C_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubaa/US10C_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	ID	Description
1	1535	100.0	296	10	US-09-941-947A-34
2	1535	100.0	296	14	Sequence 34, Appl
3	685.5	44.7	303	9	US-10-218-118-10
4	685.5	44.7	303	9	Sequence 3, Appl
5	367.5	23.9	225	16	US-09-920-923-3
6	367.5	23.9	225	16	Sequence 133, App
7	310.5	20.2	410	9	US-09-847-081B-2
8	310.5	20.2	410	9	Sequence 2, Appli
9	302.5	19.7	410	16	US-10-038-854-130
10	302.5	19.7	410	16	Sequence 130, App
11	300.5	19.6	412	9	US-10-038-854-131
12	300.5	19.6	412	9	Sequence 131, App
13	270.5	17.6	402	15	US-10-371-307-76
14	232	15.1	342	14	Sequence 76, Appl
15	213	13.9	314	14	Sequence 318, App
16	213	13.9	314	14	Sequence 8563, Ap
17	213	13.9	314	14	Sequence 9192, Ap
18	213	13.9	314	14	Sequence 16, Appl

16	201.5	13.1	1268	15	US-10-438-784-3	Sequence 3, Appli
17	196.5	12.8	275	15	US-10-438-784-7	Sequence 7, Appli
18	178	11.6	292	9	US-09-738-626-6173	Sequence 6173, Ap
19	172	11.2	374	9	US-09-820-004-2	Sequence 2, Appli
20	171.5	11.2	417	9	US-09-820-004-4	Sequence 4, Appli
21	171.5	11.2	417	9	US-09-820-004-5	Sequence 5, Appli
22	171.5	11.2	417	9	US-09-820-004-6	Sequence 6, Appli
23	168.5	11.0	363	14	US-10-358-917-2	Sequence 2, Appli
24	164.5	10.7	287	14	US-10-358-917-10	Sequence 10, Appli
25	164	10.7	279	16	US-10-038-854-16	Sequence 16, Appl
26	163	10.6	390	15	US-10-369-493-18574	Sequence 18574, A
27	159.5	10.4	416	14	US-10-205-194-67	Sequence 67, Appl
28	155.5	10.1	255	16	US-10-038-854-128	Sequence 128, App
29	143.5	9.3	520	14	US-10-128-714-3505	Sequence 3505, Ap
30	143.5	9.3	528	14	US-10-128-714-8505	Sequence 8505, Ap
31	141.5	9.2	354	15	US-10-369-493-13039	Sequence 13039, A
32	140	9.1	304	9	US-09-738-626-4195	Sequence 4195, Ap
33	139.5	9.1	448	14	US-10-032-585-7355	Sequence 7355, Ap
34	134	8.7	254	16	US-10-038-854-129	Sequence 129, App
35	131.5	8.6	300	14	US-10-156-761-9193	Sequence 9193, Ap
36	130.5	8.5	460	15	US-10-369-493-2381	Sequence 2381, Ap
37	126.5	8.2	413	15	US-10-429-949-5	Sequence 5, Appli
38	124	8.1	444	15	US-10-369-493-22063	Sequence 22063, A
39	121.5	7.9	356	15	US-10-369-493-3321	Sequence 3321, Ap
40	118	7.7	194	16	US-10-038-854-356	Sequence 356, App
41	116	7.6	388	14	US-10-024-130A-6	Sequence 6, Appli
42	116	7.6	410	14	US-10-024-130A-2	Sequence 2, Appli
43	111	7.2	38	16	US-10-038-854-132	Sequence 132, App
44	111	7.2	328	14	US-10-156-761-11454	Sequence 11454, A
45	109	7.1	131	14	US-10-149-759-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1

US-09-941-947A-34
; Sequence 34, Application US/09941947A
; Publication No: US2003003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picatoggio, Steve E.
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: C11903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 34
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-34

Query Match 100.0%; Score 1535; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.7e-160;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDVDDQTLGTFHADQPSQMPERLQQ 60

Db 1 MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDVDDQTLGTFHADQPSQMPERLQQ 60

Qy 61 LEMKTRQYAGSQMHFAFAQVAVAHADIAPIAYADHLEGFAMDVRETRYLTLDITLR 120

b 61 LEMKTRQAVAGSOMHEPFAAFQEVAMAHDIAPAFDHEGFANDVRETRYLTLDLTLR 120
y 121 YCHVAVGVGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
b 121 YCHVAVGVGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
y 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
b 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
y 241 KIGVKVEQAGKQAWDRHQSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
b 241 KIGVKVEQAGKQAWDRHQSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 2

S-10-218-118-10

Sequence 10, Application US/10218118

Publication No. US20030148319A1

GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia

APPLICANT: Rowliere, Pierre

APPLICANT: Picatagallo, Stephen

APPLICANT: Cheng, Qiong

TITLE OF INVENTION: Genes Encoding Carotenoid Compounds

FILE REFERENCE: CL1876 US NA

CURRENT APPLICATION NUMBER: US/10/218,118

PRIOR FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: 60/312,646

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Microsoft Office 97

SEQ ID NO 10

LENGTH: 296

TYPE: PRT

ORGANISM: Pantoea stewartii

S-10-218-118-10

Query Match 100.0%; Score 1535; DB 14; Length 296;

Best Local Similarity 100.0%; Pred. No. 2.7e-160;

Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 MAVGSKSFATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSSQMPQRLOQ 60

b 1 MAVGSKSFATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSSQMPQRLOQ 60

y 61 LEMKTRQAVAGSOMHEPFAAFQEVAMAHDIAPAFDHEGFANDVRETRYLTLDLTLR 120

b 61 LEMKTRQAVAGSOMHEPFAAFQEVAMAHDIAPAFDHEGFANDVRETRYLTLDLTLR 120

y 121 YCHVAVGVGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

b 121 YCHVAVGVGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

y 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVSSMAGLAQLPLRSAMAIATAKQVYR 240

b 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVSSMAGLAQLPLRSAMAIATAKQVYR 240

y 241 KIGVKVEQAGKQAWDRHQSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

b 241 KIGVKVEQAGKQAWDRHQSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 3

S-09-547-267-3

Sequence 3, Application US/09547267

Patent No. US20020147371A1

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Agolpous

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/547,267

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 303 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-547-267-3

Query Match 44.7%; Score 686.5; DB 9; Length 303;

Best Local Similarity 48.7%; Pred. No. 7.8e-67;

Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;

Qy 1 MAVGSKSFATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFHADQPSSQMPQRLOQ 60

Db 12 IAQGSQSPAQAALPPGIREDTVMYAWCRHADDVDDQVWGSAPEAGGD--POARLGA 69

Qy 61 LEMKTRQAVAGSOMHE-----PAPAAQEVAMAHDIAPAFDHEGFANDVRETRYLT 115

Db 70 LRADTLAA-----LHEDGPMSPFAALRQVARRHDFPDLWPMDLIEGFAMDVADREYRSL 124

Qy 116 DDTLRYCYHVAGVVGVLMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYL 175

Db 125 DDVLEYSYHVAGVVGVMARVMGVQDDAVLDRACDLGLAFQLTNIARDIVDDAIGRCYL 184

Qy 176 PESWLEEBGLTKANYAAPENROALSRIAGRLVREABPYVSSMAGLAQLPLRSAMAIATA 235

Db 185 PADWLAEG---ATVEGPPVSDALYSVIRLLDAAEPPYASARQGLPHLPRAWSIAAA 241

Qy 236 KQYRKIGYKVGQAGKQAWDRHQSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 295

Db 242 LRIYRAIGTRIGQGGPEATQRISTSKAKIGLLARGGLDAAASRLRGHSRDLGWTREP 301

RESULT 4

US-09-920-923-3

Sequence 3, Application US/09920923

Publication No. US2003002273A1

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Taygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/09/920,923

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 303
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
-09-920-923-3

Query Match 44.7%; Score 686.5; DB 10; Length 303;
Best Local Similarity 48.7%; Pred. No. 7.9e-67;
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;
1 MAVSKSFATSTLFDKATRSVLMVWCRHCDVDDVTLGFHADQPSQMPQRLQQ 60
12 IAQSOSFQAQAKLMPGIREDTNLYWCRHDDVDDVQVMSAPFAGGD--PQARLGA 69
61 LEMTKQAYAGSOMHE-----PAPAFAQFQVAMAHDIAPAYAFDEHGFAMVRETRYLTL 115
70 LRADTLAA-----LHEDGPMSPPPAALRQVARRHDPDLAPMDLIEGFAMVADREYRSI 124
116 DDTLYRYCYHVAGVGLMAQIMGVDRNATLGRACDGLAFOLTNIAARDIVDDAQVGRCYL 175
125 DDVLEYSTHVAGVGVWMAKVMGVQDDAVLGRACDGLAFOLTNIAARDIVDDAAIGRCYL 184
176 PESMLEBEGLTKANYAAPENRQALSRAGLRVREAPYVSSMAGLAQLPLRSAMAIATA 235
185 PADMLAERAG--ATVEGVPVSALYSVIRLLDAARPYASARQGLPHLPCCAWSIAAA 241
236 KQVTRKIGVKVQKQAKQAWDHQRSTSTAEKLTLLTASGQAVTSRMKTYPPRPAHLQRP 295
242 LEIYRAIGTRIQGGPEAYRQRTSTSKAAKIGLLARGGLDAAASRLRGCEISRDGLWTRP 301

SULT 5

-10-038-854-133
Sequence 133, Application US/10038854
Publication No. US2004002278A1

GENERAL INFORMATION:

APPLICANT: Spytex, Kimberly A
APPLICANT: Li, Li
APPLICANT: Wolenc, Adam R
APPLICANT: Vernet, Corine
APPLICANT: Eisen, Andrew J
APPLICANT: Liu, Xiaohong
APPLICANT: Malyankar, Uziel M
APPLICANT: Shimkets, Richard A
APPLICANT: Tchernev, Velizar
APPLICANT: Spaderna, Steven K
APPLICANT: Gorman, Linda
APPLICANT: Kekuda, Ramesh
APPLICANT: Patturajan, Meera
APPLICANT: Gusev, Vladimir Y
APPLICANT: Gangolli, Esha A
APPLICANT: Guo, Xiaojia S
APPLICANT: Shenoy, Suresh G
APPLICANT: Rastelli, Luca
APPLICANT: Casman, Stacie J
APPLICANT: Soldog, Ferenc
APPLICANT: Burgess, Catherine E
APPLICANT: Edinger, Shlomit R
APPLICANT: Ellerman, Karen
APPLICANT: Gunther, Erik
APPLICANT: Smithson, Glenna
APPLICANT: Millet, Isabelle
APPLICANT: MacDougall, John R

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-230

CURRENT APPLICATION NUMBER: US/10/038,854

CURRENT FILING DATE: 2003-01-22

PRIOR APPLICATION NUMBER: 60/259,928

PRIOR FILING DATE: 2000-12-29

PRIOR APPLICATION NUMBER: 60/259,415
PRIOR FILING DATE: 2001-01-02
PRIOR APPLICATION NUMBER: 60/259,785
PRIOR FILING DATE: 2001-01-04
PRIOR APPLICATION NUMBER: 60/269,814
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/279,832
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/279,833
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/279,863
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/283,889
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/284,447
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/286,683
PRIOR FILING DATE: 2001-04-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 411
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 133
LENGTH: 225
TYPE: PRT
ORGANISM: Homo sapiens
US-10-038-854-133

Query Match 23.9%; Score 367.5; DB 16; Length 225;
Best Local Similarity 46.4%; Pred. No. 6.8e-32;
Matches 85; Conservative 22; Mismatches 57; Indels 19; Gaps 5;

QY 102 GFMDVR-----STRYLTDDTLRYCYHVAGVGLMAQIMGVDR-----NATLD 146
DB 43 GWAMDLEKREKNLQRYATFEDLLRYCYVAGTVGLMWARLMGVKLEDPADWOLEVLD 102

QY 147 -RACDGLFOLTNIAARDIVDDAQVGRG--YLPESWLEBEGLTANYAAPENR-QALSRI 202
DB 103 LRACDGLALQLTNIARDVGEDARSGCRVYLPTEWLSQYGLSLEDLLAPENTDKRIRRV 162

QY 203 AGRVREABPYVSSMAGLAQLPLRSAMAIATAKQVTRKIGVKVQKQAKQAWDHQRSTST 262
DB 163 LRRLDNARAYVEDALTGLAGLPQSRFPFIAAAPQVYAGIGDAIBANGYDVFRRRAKTRK 222

QY 263 AEK 265

DB 223 GEK 225

RESULT 6

US-09-847-081B-2
Sequence 2, Application US/09847081B
Patent No. US2002012846A1
GENERAL INFORMATION:
APPLICANT: BAYER AG
TITLE OF INVENTION: DNA encoding the tobacco phytoene synthase
FILE REFERENCE: Le A 34 326
CURRENT APPLICATION NUMBER: US/09/847,081B
CURRENT FILING DATE: 2001-05-02
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 440
TYPE: PRT
ORGANISM: Nicotiana tabacum
US-09-847-081B-2

Query Match 21.9%; Score 336.5; DB 9; Length 440;
Best Local Similarity 30.3%; Pred. No. 4.7e-28;
Matches 88; Conservative 52; Mismatches 131; Indels 19; Gaps 6;

QY 5 SKSFATSTLFDKATRSVLMVWCRHCDVDDVTLGFHADQPSQMPQRLQQLEMK 64
DB 158 AKTYLGTCLMTTPERRAIWAIYVWCRTDSELVDGPN-----ASHITPQALDRWETR 209

Y 65 TRQAYAGSQHPEPAPAAFOEVAMAH--DIAPAYAFDHLEGPAMDVRETRYLTLDLTLRYC 122
b 210 LEDFSG-RPFDMLDAALSDTVSRPVDIOPFR--DMIEGEMMDLWKSRYKTFDELYLC 266
Y 123 YHVAGVVGGLMAQIMGV--RDNATLD----RACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
b 267 YTVAGTVGLNSVPVAGIAPESKATTESVYNAALALGLANQLTNILRDVGEDARGRVYLP 326
Y 177 ESWLEBEGLTKANYAAPENRQALSRIAGRLVREABPYTVSSMAGLAQLPLRSANAIATAK 236
b 327 QDELAQAGLSDEDFIAGRVTDKRNFMKKQIQARCKPDESEKGVTELDLSASRPVLITAL 386
Y 237 QVYRKIGVKGVEAGKQAMDHROSTSTAELKLTLLLTASQVATSRMKTYP 286
b 387 LLYRKILDEIANDYNNFTFRAYVSKPKLLTLPAYAKSLVPPNRTSSP 436

RESULT 7

S-09-847-081B-4
Sequence 4, Application US/09847081B
Patent No. US20020128464A1
GENERAL INFORMATION:

APPLICANT: BAYER AG
TITLE OF INVENTION: DNA encoding the tobacco phytoene synthase

FILE REFERENCE: Le A 34 326

CURRENT APPLICATION NUMBER: US/09/847,081B

CURRENT FILING DATE: 2001-05-02

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 410

TYPE: PRT

ORGANISM: Nicotiana tabacum

FEATURE:

NAME/KEY: unsure

LOCATION: 135, 139

OTHER INFORMATION: Xaa is unknown or other

S-09-847-081B-4

Query Match 20.2%; Score 310.5; DB 9; Length 410;
Best Local Similarity 28.2%; Pred. No. 3.1e-25;
Matches 81; Conservative 56; Mismatches 131; Indels 19; Gaps 5;

Y 5 SKSFATASTLFDKTRRSVLMYAWCRHCDVDIDQTLGFGHADQPSQMPQRLQOLEMK 64
b 130 AKTFYGTXTMTPEERRAIWAIYVWCERTDELVDGPN-----ASHITPQALDRWEDR 181
Y 65 TRQAYAGSQHPEPAPAAFOEVAMAH--DIAPAYAFDHLEGPAMDVRETRYLTLDLTLRYC 122
b 182 LEDVFSG-RPFDMLDAALSDTVSKFPVDIOPFR--DMIEGEMMDLWKSRYKTFDELYLC 238
Y 123 YHVAGVVGGLMAQIMGVDRN-----ATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
b 239 YTVAGTVGLNSVPVAGIAPDSKATTESVYNAALALGLANQLTNILRDVGEDARGRVYLP 298
Y 177 ESWLEBEGLTKANYAAPENRQALSRIAGRLVREABPYTVSSMAGLAQLPLRSANAIATAK 236
b 299 QDELAQAGLFDDIPDIPAGKVTDKWRSFMKKQIQARCKPDEABEGVTQLSSASRPVWASL 358
Y 237 QVYRKIGVKGVEAGKQAMDHROSTSTAELKLTLLLTASQVATSRMKT 283
b 359 LLYRQILDEIANDYNNFTFRAYVSKPKLLISLPIAYAKSLVPPNRT 405

RESULT 8

S-10-038-854-130
Sequence 130, Application US/10038854
Publication No. US20040022781A1
GENERAL INFORMATION:

APPLICANT: spytek, Kimberly A

APPLICANT: Li, Li

APPLICANT: Wolenc, Adam R

APPLICANT: Vernet, Corine
APPLICANT: Bisen, Andrew J
APPLICANT: Liu, Xiaohong
APPLICANT: Malvankar, Uriel M
APPLICANT: Shinkov, Richard A
APPLICANT: Tchernev, Velizar
APPLICANT: Spaderna, Steven K
APPLICANT: Gorman, Linda
APPLICANT: Kekuda, Ramesh
APPLICANT: Patturajan, Meera
APPLICANT: Gusev, Vladimir Y
APPLICANT: Gangolli, Bsha A
APPLICANT: Guo, Xiaojia S
APPLICANT: Shenoy, Suresh G
APPLICANT: Rastelli, Luca
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc
APPLICANT: Burgess, Catherine E
APPLICANT: Edinger, Shlomit R
APPLICANT: Ellerman, Karen
APPLICANT: Gunther, Erik
APPLICANT: Smithson, Glennda
APPLICANT: Millet, Isabelle
APPLICANT: MacDougall, John R
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-230
CURRENT APPLICATION NUMBER: US/10/038,854

CURRENT FILING DATE: 2003-01-22

PRIOR APPLICATION NUMBER: 60/258,928

PRIOR FILING DATE: 2000-12-29

PRIOR APPLICATION NUMBER: 60/259,415

PRIOR FILING DATE: 2001-01-02

PRIOR APPLICATION NUMBER: 60/259,785

PRIOR FILING DATE: 2001-01-04

PRIOR APPLICATION NUMBER: 60/269,814

PRIOR FILING DATE: 2001-02-20

PRIOR APPLICATION NUMBER: 60/279,832

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/279,833

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/279,863

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/283,889

PRIOR FILING DATE: 2001-04-13

PRIOR APPLICATION NUMBER: 60/284,447

PRIOR FILING DATE: 2001-04-18

PRIOR APPLICATION NUMBER: 60/286,683

PRIOR FILING DATE: 2001-04-25

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 411

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 130

LENGTH: 436

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-10-038-854-130

Query Match 20.2%; Score 310.5; DB 16; Length 436;
Best Local Similarity 27.8%; Pred. No. 3.4e-25;
Matches 80; Conservative 58; Mismatches 131; Indels 19; Gaps 5;

QY 5 SKSFATASTLFDKTRRSVLMYAWCRHCDVDIDQTLGFGHADQPSQMPQRLQOLEMK 64
b 156 AKTFYGTXTMTPEERRAIWAIYVWCERTDELVDGPN-----ASHITPQALDRWESR 207
QY 65 TRQAYAGSQHPEPAPAAFOEVAMAH--DIAPAYAFDHLEGPAMDVRETRYLTLDLTLRYC 122
b 208 LEDLFRG-RPFDMLDAALSDTVTKFPVDIOPFR--DMIEGEMMDLWKSRYKTFDELYLC 264
QY 123 YHVAGVVGGLMAQIMGVDRN-----ATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
b 265 YTVAGTVGLNSVPVAGIAPDSKATTESVYNAALALGLANQLTNILRDVGEDARGRVYLP 324

362 LYRKILDEIANDYNNFTKRAYVSKSKLLIALPIAYAKSLVPPTKT 407

3-10-401-321-76
Sequence 76, Application US/10401321
Publication No. US20030233679A1
GENERAL INFORMATION:

APPLICANT: Brown, Sherri M.
APPLICANT: Heck, Gregory R.
APPLICANT: Pillier, Kenneth J.
APPLICANT: Kilshore, Ganesh M.
APPLICANT: Ellich, Tedd D.
APPLICANT: Logusch, Eugene W.
APPLICANT: Rao, Sudabathula
APPLICANT: Ream, Joel E.
APPLICANT: Logusch, Sherry J.
APPLICANT: Baerson, Scott R.

TITLE OF INVENTION: Methods for Controlling Gibberellin Levels
FILE REFERENCE: 11899.0216.DVUS01 (MOBT:218--1)
CURRENT APPLICATION NUMBER: US/10/401,321
CURRENT FILING DATE: 2003-03-27

NUMBER OF SEQ ID NOS: 89
SOFTWARE: PatentIn version 3.2

SEQ ID NO 76
LENGTH: 412
TYPE: PRT
ORGANISM: Lycopersicon esculentum
S-10-401-321-76

Query Match 19.6%; Score 300.5; DB 15; Length 412;
Best Local Similarity 29.0%; Pred. No. 4e-24;
Matches 83; Conservative 51; Mismatches 135; Indels 17; Gaps 5;

Y 5 SKSFATASTLFDKTRRSVLMYVWCHDDVDDQTLGFHADQSSQMPERLOQLEMK 64
b 132 AKTFNLGTMLTPERRAIWAIYVWCRTEDELVDGPN-----ASYITPAALORWNR 183
Y 65 TRQVAGSQMHPEAPAFQFVA-WAHDIAFAYAFDELEGFAMDVRETRYLTLDLTYCY 123
b 184 LEDVFNRPFTMDLGSALSDTVSNFPVDIQPFR--DMIEGNEXDLRSRYKNFDELYLYCY 241
Y 124 HVAGWGLMAQINGV--RDNATLD-----RACDLGLAFQLTNIARDIVDDAQVGRCLPE 177
b 242 YVAGTVGLMSVPINGIAPESKATITESVYNAALALGIANQLTNILRDVGEDARRGRVLPQ 301
Y 178 SWLEEGHGTANYAAPENROALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAQ 237
b 302 DELAQAGLSDSIDIFAGRVTDKRWIFPMKKQIHRARKFDEAEKGVTELSASRPVWASLV 361
Y 238 VYRKIGVKEQAGKQADHRQSTSTAELKTLTLLTASQAVTSRMKT 283
b 362 LYRKILDEIANDYNNFTKRAYVSKSKLLIALPIAYAKSLVPPTKT 407

RESULT 12
US-10-259-194A-318
Sequence 318, Application US/10259194A
Publication No. US20040010815A1
GENERAL INFORMATION:

APPLICANT: Lange, Markus B.
APPLICANT: Ghassemian, Majid
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazebrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Moughamer, Todd
APPLICANT: Provart, Nicholas
APPLICANT: Ricke, Darrell
APPLICANT: Zhu, Tong

TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

FILE REFERENCE: 70029-NP
CURRENT APPLICATION NUMBER: US/10/259,194A
CURRENT FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,743
PRIOR FILING DATE: 2002-04-04
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04

NUMBER OF SEQ ID NOS: 662
SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
SEQ ID NO 318
LENGTH: 402

ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: X-region
LOCATION: (58)..(58)
OTHER INFORMATION: Xaa = any naturally occurring amino acid
US-10-259-194A-318

Query Match 17.6%; Score 270.5; DB 15; Length 402;
Best Local Similarity 27.2%; Pred. No. 7.7e-21;
Matches 78; Conservative 60; Mismatches 126; Indels 23; Gaps 8;

QY 5 SKSFATASTLFDKTRRSVLMY----AWCRHDDVIDDQTLGFHADQSSQMPERLOQ 60
Db 126 AKTFYLETQLTMPERRKAVWAIYGMVLWCRTEDELVDGPN-----SSYITPKALDR 177
QY 61 LEMKTRQVAGSQMHPEAPAFQFVAWAMHDIAFAYAF-DHLEGFAMDVRETRYLTLDLTL 119
Db 178 WEKRLDLEFG-RPYDMYDAALSDTVSKFPVDIQPFPKDMIEGMRLDLWKSRYRSFDELY 236
QY 120 RYCYHVAGVYGLMAQINGVRDN-----ATLDACDLGLAFQLTNIARDIVDDAQVGRCL 173
Db 237 LYCYVAGTVGLMTVPWGIAPDSKASTESVYNAALALGIANQLTNILRDVGEEER-GRI 295
QY 174 YEPESWLEEGHGTANYAAPENROALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAI 232
Db 296 YLPLDELAEEAGLFEEDIFRGKVTWKWKMGQILR-ARLFFDEAEKGVHLDLSASRPV 354
QY 233 ATAKQVYRKIGVKEQAGKQADHRQSTSTAELKTLTLLTASQAVTS 279
Db 355 LASLWLYRQILDALEANDYNNFTKRAYVWAKKLLSLPVAYARAAYA 401

RESULT 13
US-10-156-761-8563
Sequence 8563, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, WASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8563
LENGTH: 342
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8563

Query Match 15.1%; Score 232; DB 14; Length 342;
Best Local Similarity 27.6%; Pred. No. 1.1e-16;
Matches 90; Conservative 40; Mismatches 136; Indels 60; Gaps 12;

6 KSPATATLFDKTRRSVLMYAWCRHCDVID--DQTLGFHADQPSQMPQOR---LQO 60
32 KTYFLATRLLPVRRPAVHALYGFARWADDIVDSLOTTVG-----PAVRSALGR 81

61 LEMKTRQAYAGSQHPEPAFAQFVAMAHDIAPAYAFDHLEGFAMDVRETRYLTLDLTLR 120
82 LQESLRTGLRBGHSFVPLALASTARAYADHQHFGDFMTAMRSDLVGTGYETADLRG 141
121 YCYRVAGVVGGLMAQIMGV--RDNATLDRACDLGLAFQLTNIARDIVDDAQVRCYLPE 177
142 YMHGSAVIGLQMLPVLTGVVPREAA-PRHAAAGVAFQLTNFIIRDVGEDLDRGVILPA 200

178 SWLEEEGLTK--ANYAAPENR-----QALSRIAG---RLVRAEYVYSSMAGLAQLPL 226
201 DLLRAHGVDRLLRSDTGRDRRIITRALKAVEGLTRGVYREAAP-----GLAWLDP 253
227 RSAWAIATAKOVYRKIGVKVEQAGKQAWDHROSTSTASK-----LTLTLTASQAV 277
254 VARPCTRTAFVLYGGILDADDDG-YAVVHRAVVPERRRAAVALDGLVRLGVARLGSRA 312
278 TSMKTYP-----PREAH 290
313 DARLPTLPDWRPLALVPRSSPREAH 338

SULT 14
-10-156-761-9192
Sequence 9192, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156.761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9192
LENGTH: 316
TYPE: PRT
ORGANISM: Streptomyces avermitilis
-10-156-761-9192

Query Match 15.0%; Score 229.5; DB 14; Length 316;
Best Local Similarity 30.0%; Pred. No. 1.8e-16;
Matches 81; Conservative 38; Mismatches 134; Indels 17; Gaps 9;

5 SKSPATATLFDKTRRSVLMYAWCRHCDVIDDQTLGFHADQPSQMPQORLQOLEMK 64
30 ARNFAVGIRLLPTPKRRAMSAVAFPSRVDDIGG---ALAPDVKAARLEDTTR--ALLSR 84

65 TRQAYAGSQHPEPAFAQFVAMAHDIAPAYAFDHL-EGFAMDVRETRYLTLDLTLRYCY 123
85 VRDGRVDEDDTPVAVALAHAAEQFPI-PLAGLDELIDGLMDVGRGETYETWDDLKYCR 143
124 HVAGVVGGLMAQIMGVRNAT-LRACD-----LGLAFQLTNIARDIVDDAQVRCYLPE 178
144 CVAGAIGRVSLGVTFGTFCARGAERASEYADTLGLALQLTNILEDVREDDEAGGRYLPAD 203

QY 179 WLEEGTLKANYAAPENRQALSRIAGRL---VRBAEYVYSSMAGLAQLPLRSNAIATA 235
DB 204 DLAKFGCS-AGFDRPIPEG-SDPAGLVHFEVRRARALFAEGYRLLPMLDRRSACVAM 261
QY 236 KOVYRKIGVKVEQAGKQAWDHROSTSTASK 265
DB 262 AGIYRLLDRIEREPEAVLRGRVSLPGREX 291

RESULT 15
US-10-128-713A-16
Sequence 16, Application US/10128713A
Publication No. US20030170847A1
GENERAL INFORMATION:
APPLICANT: Bramucci, Michael G
TITLE OF INVENTION: Genes Involved in Isoprenoid Compound Production
FILE REFERENCE: CL-1788
CURRENT APPLICATION NUMBER: US/10/128.713A
CURRENT FILING DATE: 2002-04-22
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Microsoft Office 97
SEQ ID NO 16
LENGTH: 314
TYPE: PRT
ORGANISM: Rhodococcus erythropolis
US-10-128-713A-16

Query Match 13.9%; Score 213; DB 14; Length 314;
Best Local Similarity 27.7%; Pred. No. 1.1e-14;
Matches 83; Conservative 46; Mismatches 143; Indels 28; Gaps 8;

QY 6 KSPATATLFDKTRRSVLMYAWCRHCDVIDDQTLGFHADQPSQMPQORLQOLEMK 65
DB 20 RTYFLATRLLPERRRAVHALYAFARVDDVDEPS-GPH-ERGTVLADVERAAVTALDN 77
QY 66 RQAYAGSQHPEP-----AFAAQFVAMAHDIAPAYAFDHLEGFAMDVRETR----VLT 115
DB 78 PTATGGPFSTIPLDLTRVLPFAFADAVTFOIPRAYFDAPFESMEXDAPDTAKFRPVNTM 137
QY 116 DDTLRYCYHVAGVVGGLMAQIMGV--DNATLDRACDLGLAFQLTNIARDIVDDAQVRC 173
DB 138 DELAEYMYGSAVVGIGLQMLPIGVSVFQEAUVFASNLGEAFQLTNFIIRDVGEDLDRGL 197
QY 174 YLPSWLEEEGLTKANYAAPENRQALSRIAGRLVREAPY-----YVSSMAGLAQLPL 226
DB 198 YLPAGEFAAFGV--DIEMLSHGRTGTGVVVRKRALAHFIATVTRGRYSAESGIPMLDR 254
QY 227 RSAWAIATAKOVYRKIGVKVEQAGKQAWDHROSTSTASKLTL-----LTLTASQAVTSMK 282
DB 255 RVQPSIRTAFLVLYGAILDQVERADFRILHRRVSVPGTRLRVAAPGLVRSATYAANKMR 314

Search completed: February 29, 2004, 15:28:05
Job time : 22.6527 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model
in on: February 29, 2004, 14:35:44 ; Search time 8.99489 Seconds
(without alignments)
1698.885 Million cell updates/sec

file: US-09-941-947A-34

Effect score: 1535

quence: 1 MAVGSKSFATATLFDKTR.....VTGRMKTYPPRPAHLWQRP1 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	1368	89.1	296	1	US-07-783-705A-5	Sequence 5, Appli
2	1028	67.0	309	1	US-08-331-004A-2	Sequence 2, Appli
3	1028	67.0	309	5	PCT-US95-13937A-2	Sequence 2, Appli
4	997.5	65.0	308	1	US-08-095-726-6	Sequence 6, Appli
5	997.5	65.0	308	1	US-08-096-043-6	Sequence 6, Appli
6	997.5	65.0	308	1	US-08-093-577-6	Sequence 6, Appli
7	997.5	65.0	308	1	US-08-096-223A-6	Sequence 6, Appli
8	686.5	44.7	303	3	US-08-660-645A-3	Sequence 3, Appli
9	686.5	44.7	303	3	US-09-298-718-3	Sequence 3, Appli
10	686.5	44.7	303	3	US-09-546-969-3	Sequence 3, Appli
11	686.5	44.7	303	3	US-08-980-832-3	Sequence 3, Appli
12	686.5	44.7	303	4	US-09-547-267-3	Sequence 3, Appli
13	686.5	44.7	303	4	US-09-920-923B-3	Sequence 3, Appli
14	318.5	20.7	410	1	US-08-579-667-6	Sequence 2, Appli
15	318.5	20.6	410	1	US-08-579-667-2	Sequence 2, Appli
16	315.5	20.6	410	1	US-08-579-667-8	Sequence 8, Appli
17	311.5	20.3	413	1	US-08-579-667-4	Sequence 4, Appli
18	309.5	20.2	400	4	US-09-691-270A-14	Sequence 14, Appl
19	307.5	20.0	400	4	US-09-691-270A-27	Sequence 27, Appl
20	304.5	19.8	402	3	US-09-180-342-3	Sequence 3, Appli
21	302.5	19.7	410	4	US-09-691-270A-28	Sequence 28, Appl
22	286.5	18.7	408	4	US-09-691-270A-2	Sequence 2, Appli
23	252	16.4	252	4	US-09-691-270A-12	Sequence 12, Appl
24	249.5	16.3	248	4	US-09-691-270A-16	Sequence 16, Appl
25	243	15.8	242	4	US-09-691-270A-10	Sequence 10, Appl
26	227.5	14.8	186	4	US-09-691-270A-4	Sequence 4, Appli
27	179.5	11.7	673	4	US-09-091-725-13	Sequence 13, Appl

Query Match

89.1% Score 1368; DB 1; Length 296;

ALIGNMENTS

RESULT 1

US-07-783-705A-5

; Sequence 5, Application US/07783705A

; Patent No. 5429939

; GENERAL INFORMATION:

; APPLICANT: Misawa, No. 5429939ihiko

; APPLICANT: Kobayashi, Kazuo

; APPLICANT: Nakamura, Katsumi

; APPLICANT: Yamano, Shigeyuki

; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE

; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ladax & Party

; STREET: 26 West 61 Street

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10023

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: N/A

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07783,705A

; FILING DATE: 19911023

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 1-103078

; FILING DATE: 21-APR-1989

; APPLICATION NUMBER: JP 2-53225

; FILING DATE: 05-MAR-1990

; APPLICATION NUMBER: US 07/519,011

; FILING DATE: 19-APR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Schwadron, Janet I.

; REGISTRATION NUMBER: 33,778

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-708-1935

; TELEFAX: 212-246-5959

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 296 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-07-783-705A-5

Best Local Similarity 88.9%; Pred. No. 5.9e-143;
Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

Y 1 MAVGSKSPATSTLFDKATRSVLMYAWCRHCDVDVDDQTLGPHADQSSQMPQRLOQ 60
|||||
b 1 MAVGSKSPATSKLFDKATRSVLMYAWCRHCDVDVDDQTLGPHADQSSQMPQRLOQ 60
|||||

Y 61 LEMKTRQAYAGSOMHEPAPAAQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYTLTDDTLR 120
|||||
b 61 LEMKTRQAYAGSOMHEPAPAAQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYTLTDDTLR 120
|||||

Y 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQFLNIAARDIVDDAQVGRCYLPESWL 180
|||||
b 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQFLNIAARDIVDDAQVGRCYLPESWL 180
|||||

Y 181 EEBGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240
|||||
b 181 EEBGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240
|||||

Y 241 KIGVKVEQAGKQAWDHROSTSTAEBKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
|||||
b 241 KIGVKVEQAGKQAWDHROSTSTAEBKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
|||||

RESULT 2

S-08-331-004A-2

Sequence 2, Application US/08331004A

Patent No. 5618988

GENERAL INFORMATION:

APPLICANT: Hauptmann, Randal

APPLICANT: Eschenfeldt, William H

APPLICANT: English, Jami

APPLICANT: Brinkhaus, Friedhelm L

TITLE OF INVENTION: Enhanced Carotenoid Accumulation

TITLE OF INVENTION: in Storage Organs of Genetically

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corporation, Law Dept

STREET: 55 Shuman Boulevard, Suite 600

CITY: Naperville

STATE: IL

COUNTRY: USA

ZIP: 60563-8437

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/331,004A

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, No. 5618988val B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7087172447

TELEFAX: 7087172430

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

S-08-331-004A-2

Query Match 67.0%; Score 1028; DB 1; Length 309;

Best Local Similarity 64.9%; Pred. No. 2.7e-105;

Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

QY 61 LEMKTRQAYAGSOMHEPAPAAQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYTLTDDTLR 120
|||||
Db 74 LRTLTUAAFEQEMQDPAPAAQEVVALTHGITPRMALDHLGDFAMDVAGTRVTFEDTLR 133
|||||

QY 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQFLNIAARDIVDDAQVGRCYLPESWL 180
|||||
Db 134 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQFLNIAARDIVDDAQVGRCYLPESWL 193
|||||

QY 181 EEBGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240
|||||
Db 194 QDGLTPENYAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 253
|||||

QY 241 KIGVKVEQAGKQAWDHROSTSTAEBKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296
|||||
Db 254 EIGIKVKAAGGSANDRRQHTSKGKIAMLMAPGGVIRAKTTRVTRPAGLWQRPV 309
|||||

RESULT 3

PCT-US95-13937A-2

Sequence 2, Application PC/TUS9513937A

GENERAL INFORMATION:

APPLICANT: Hauptmann, Randal

APPLICANT: Eschenfeldt, William H

APPLICANT: English, Jami

APPLICANT: Brinkhaus, Friedhelm L

TITLE OF INVENTION: Enhanced Carotenoid Accumulation

TITLE OF INVENTION: in Storage Organs of Genetically

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corporation, Law Dept

STREET: 55 Shuman Boulevard, Suite 600

CITY: Naperville

STATE: IL

COUNTRY: USA

ZIP: 60563-8437

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13937A

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, Norval B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7087172447

TELEFAX: 7087172430

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-13937A-2

Query Match 67.0%; Score 1028; DB 5; Length 309;

Best Local Similarity 64.9%; Pred. No. 2.7e-105;

Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

QY 1 MAVGSKSPATSTLFDKATRSVLMYAWCRHCDVDVDDQTLGPHADQSSQMPQRLOQ 60
|||||
Db 14 MAVGSKSPATAKLFDPATRSVLMYAWCRHCDVDVDDQTLGPHADQSSQMPQRLOQ 73
|||||

QY 61 LEMKTRQAYAGSOMHEPAPAAQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYTLTDDTLR 120
|||||
Db 74 LRTLTUAAFEQEMQDPAPAAQEVVALTHGITPRMALDHLGDFAMDVAGTRVTFEDTLR 133
|||||

QY 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQFLNIAARDIVDDAQVGRCYLPESWL 180
|||||

181	EEEGHUKYAAAPENRQALSRIGLUREHPYVSSMAGLAOLPIRSAWAIAITAKQVTR	240	
QY	194	QDAGLAPENYAARENRPALARWR-RLTDAAPYIYSQAGLHDLRRSAWIAITARSVTR	252
DB	241	KIGKVQEQAGQANDHQSTSTAEKLTLLLTASQAVTSMKTYTPRPAHLWORI	296
QY	253	EIGIKVRAAGSGAWDRCHTSKGEKLAUMLAAQGVTRAKTTRVTPRAGLWOPV	308
DB			

RESULT 5

US-08-096-043-6
? Sequence 6, Application US/08096043
? Patent No. 5530189
? GENERAL INFORMATION:
? APPLICANT: Ausich, Rodney L
? APPLICANT: Brinkhaus, Friedrich L
? APPLICANT: Mukharji, Indrani
? APPLICANT: Proffitt, John H
? APPLICANT: Vargser, James G
? APPLICANT: Yen, Hwei-Che B
? TITLE OF INVENTION: Lycopene Biosynthesis in
? TITLE OF INVENTION: Genetically Engineered Hosts
? NUMBER OF SEQUENCES: 70
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Amoco Corp., Parents and Licensing Dept
? STREET: 200 E Randolph St
?

024241, 200 S Randolph St
 City: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60680-0703
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096,043
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/785,568
 FILING DATE: 30-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Galloway, No. 5530189val B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3128567180
 TELEFAX: 3128564972
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 308 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-096-043-6

Query Match

[illegible]

b 194 QDAGLAPENYAARENRPALARWR-BLIDAAEPPYISSOAGLHDLRRSAAWAIATARSYR 252
y 241 KIGVKVEQAGKAWDHROSTSTAEKLTLLTASGQAVTSRMKTYPPRPAHLWQRP1 296
b 253 EIGIKVKAAGGSANDRRQHTSKGKIAMLMAPGGVIRAKTRVTPRPAHLWQRPV 308

RESULT 6

3-08-093-577-6
Sequence 6, Application US/08093577
Patent No. 5545816

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.
APPLICANT: Brinkhaus, Friedhelm L.
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H.
APPLICANT: Yarger, James G.
APPLICANT: Yen, Huel-Che B.
TITLE OF INVENTION: Phytoene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093,577
FILING DATE: 19-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,569
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5545816val B
TELEPHONE: 3128567180
TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 308 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

S-08-093-577-6

Query Match 65.0%; Score 997.5; DB 1; Length 308;

Best Local Similarity 63.9%; Pred. No. 6.3e-102;

Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

y 1 MAVGKSFATATLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 60
b 14 MANGKSFATAKLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 73
y 61 LEWTKQVAGSQMHPAPAFQFVAMAHDIAPAYAFHLEGFAMDVETRYLIDDTLR 120
b 74 LRTLTLAAFEAGMDQAPAFQFVAMAHDIAPAYAFHLEGFAMDVETRYLIDDTLR 133
y 121 YCVHVGAVGLMAQIMGVDRNATIDRACDLGLAFQLTNIARDIVDDAQVRCYLPESWL 180
b 134 YCVHVGAVGLMAQIMGVDRNATIDRACDLGLAFQLTNIARDIVDDAQVRCYLPESWL 193
y 181 EREGTKYANAPENKQALSRAGRLVREAEFYVSSMAGLAQLPLGSAAWAIATAKOYR 240
b 194 QDAGLAPENYAARENRPALARWR-BLIDAAEPPYISSOAGLHDLRRSAAWAIATARSYR 252

Qy 241 KIGVKVEQAGKAWDHROSTSTAEKLTLLTASGQAVTSRMKTYPPRPAHLWQRP1 296
Db 253 EIGIKVKAAGGSANDRRQHTSKGKIAMLMAPGGVIRAKTRVTPRPAHLWQRPV 308

RESULT 7

US-08-096-623A-6
Sequence 6, Application US/08096623A
Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.
APPLICANT: Brinkhaus, Friedhelm L.
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H.
APPLICANT: Yarger, James G.
APPLICANT: Yen, Huel-Che B.
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Ganson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-096-623A-6

Query Match 65.0%; Score 997.5; DB 1; Length 308;

Best Local Similarity 63.9%; Pred. No. 6.3e-102;

Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

Qy 1 MAVGKSFATATLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 60
Db 14 MANGKSFATAKLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 73

116 DDTLYCYHVAGVGLMAQINGVRDNATLDRACDGLGLAFQLTNIARDIVDDAQVGRCTL 175
125 DDVLEYSYHVAGVGVWMAVWVQDDAVLDRACDGLGLAFQLTNIARDIVDDAAIGRCYL 184
176 PESWLEEGTLKANYAAPENRQALSRIAGRLVREAPFYVSSNAGLAQLPLRSAAIATA 235
185 PADWLAEG--ATVEGPVPSDALYSVIIRLLDAAEPYASARQGLPHLPFRCAWSIAAA 241
236 KOVYRKIGVKEQAGKQAWDHROSTSTAEBKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295
242 LRIYRAIGTRIRQGGPEAYRQISTSKAAKIGLLARGGLDAAASRLRGGEISRDLGLWTRP 301

RESULT 10

US-09-546-969-3

Sequence 3, Application US/09546969

Patent No. 6207409

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Passamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/546,969

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 303 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein

US-09-546-969-3

Query Match 44.7%; Score 686.5; DB 3; Length 303;

Best Local Similarity 48.7%; Pred. No. 1.6e-67;

Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;

1 MAVGSKSPATASTLFDKATRSVLMYAWCRHCDVDVDDQTLGFHADQSSQMPQRLOQ 60
12 IAQGSQSFQAQAKLMPGIREDTVLMYAWCRHADDVDDQVMSAPEAGGD--PQARLGA 69
61 LEMKTRQAYAGSQMHE-----PAPAFQEVAMAHDIAPAVAFHLEGFAMDVRETRYLTL 115
70 LRADTLAA-----LHEDGPMSPFPFALRQVARRHDFDLWPMDLIEGFAMDVADREYRSL 124
116 DDTLYCYHVAGVGLMAQINGVRDNATLDRACDGLGLAFQLTNIARDIVDDAQVGRCTL 175

125 DDVLEYSYHVAGVGVWMAVWVQDDAVLDRACDGLGLAFQLTNIARDIVDDAAIGRCYL 184
176 PESWLEEGTLKANYAAPENRQALSRIAGRLVREAPFYVSSNAGLAQLPLRSAAIATA 235
185 PADWLAEG--ATVEGPVPSDALYSVIIRLLDAAEPYASARQGLPHLPFRCAWSIAAA 241
236 KOVYRKIGVKEQAGKQAWDHROSTSTAEBKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295
242 LRIYRAIGTRIRQGGPEAYRQISTSKAAKIGLLARGGLDAAASRLRGGEISRDLGLWTRP 301

RESULT 11

US-08-980-832-3

Sequence 3, Application US/08980832B

Patent No. 6291204

GENERAL INFORMATION:

APPLICANT: Passamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/08/980,832B

CURRENT FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3

LENGTH: 303

TYPE: PRT

ORGANISM: Flavobacterium sp. R1534

US-08-980-832-3

Query Match 44.7%; Score 686.5; DB 3; Length 303;

Best Local Similarity 48.7%; Pred. No. 1.6e-67;

Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;

1 MAVGSKSPATASTLFDKATRSVLMYAWCRHCDVDVDDQTLGFHADQSSQMPQRLOQ 60
12 IAQGSQSFQAQAKLMPGIREDTVLMYAWCRHADDVDDQVMSAPEAGGD--PQARLGA 69

61 LEMKTRQAYAGSQMHE-----PAPAFQEVAMAHDIAPAVAFHLEGFAMDVRETRYLTL 115
70 LRADTLAA-----LHEDGPMSPFPFALRQVARRHDFDLWPMDLIEGFAMDVADREYRSL 124
116 DDTLYCYHVAGVGLMAQINGVRDNATLDRACDGLGLAFQLTNIARDIVDDAQVGRCTL 175
125 DDVLEYSYHVAGVGVWMAVWVQDDAVLDRACDGLGLAFQLTNIARDIVDDAAIGRCYL 184
176 PESWLEEGTLKANYAAPENRQALSRIAGRLVREAPFYVSSNAGLAQLPLRSAAIATA 235
185 PADWLAEG--ATVEGPVPSDALYSVIIRLLDAAEPYASARQGLPHLPFRCAWSIAAA 241
236 KOVYRKIGVKEQAGKQAWDHROSTSTAEBKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295
242 LRIYRAIGTRIRQGGPEAYRQISTSKAAKIGLLARGGLDAAASRLRGGEISRDLGLWTRP 301

RESULT 12

US-09-547-267-3

Sequence 3, Application US/09547267

Patent No. 6613543

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Passamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

```
62 ALEALFTLAGRMEDYVRLLPVKFYLCWESGKTLDYANDSFELEAQITQFNPRDVEGY 121
127 QRFVLYSKNLCETEAAGFAKG---LDGFMDLLKFGYPLRSLLSFDVFRSMDQVRRFIS 183
122 RFLAYSAQVQF---GYRLGVSVPFLSFRDMLR-AGP--QLLKQAWQSVYQSVRFIE 175
184 DPKLVEILNYFIKVVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELG 243
176 DEHLRQAFSPHSLVGVGNPFTTSIYTLIHALEREWGVPFGGTGALVNGMWKLTDLG 235
244 VEIRLDAEVSIOQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLRS-PASELK-- 300
236 GEIHLNARVEELVADNRVSVRLADGRIFETDVAASNADVNTYKLLGTIPVQGRRA 295
301 KQVFPSCGLVHLGVDRLYPQLAHNPFYSDPHREHDAVFKSHRLSDDDPTIYLVAP 360
296 RLERKSMNSLFLVLYFGLNQPHSLAHTTICFGPRYRELIDEIFTGSALADDFSLYHSP 355
361 CKTDPQAPACCEIKILPHIPHLDPDKLLTAEDYSALRERVVLKLERGLTDLRQHI 420
356 CVTDPSTLAPPCCASFYVLAPVPHLGNAPLQWAGPKLRIDFYDEERYMPGLRSOLVT 415
421 BEYMTPLDIQAKYYSNCGSIY-----GVVADRFRKNLGFAPQRSSELSNLYFVGGSV 472
416 QRIFTQTSRHWAIILGSLFIEPPSLTQGLFA-----ANATRSNLYLVAAGT 464
473 NPGGMPMVTLSGQVLRDKIVADLQ 497
465 HPGAGIPGWGLAESTASLTIEDLQ 489
```

RESULT 7

-08-096-623A-8

Sequence 8, Application US/08096623A

Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.
APPLICANT: Brinkhaus, Friedhelm L.
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H.
APPLICANT: Yarger, James G.
APPLICANT: Yen, Hwei-Chue B.
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551

FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-623A-8

Query Match

27.0%; Score 702.5; DB 1; Length 489;

Best Local Similarity 32.9%; Pred. No. 1.9e-63;

Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

```
QY 7 QRVIVIGAGLGLSAAISLATAGFSVQLIRKNDKVGKLNIMTKDGTDFDLGSPSILTMPH 66
DB 2 KKTVVIGAGFGGLALAIRLQAAGIPTVLLRQDKPKGGRAYVWHQDQGTTFDAGPTVITDPT 61
QY 67 IFEALFTGAGNMADYVOIQKVEPHNRNFFDGDVIDLCEDAETQRRELKLGPGTVAQF 126
DB 62 ALEALFTLAGRMEDYVRLLPVKFYLCWESGKTLDYANDSFELEAQITQFNPRDVEGY 121
QY 127 QRFVLYSKNLCETEAAGFAKG---LDGFMDLLKFGYPLRSLLSFDVFRSMDQVRRFIS 183
DB 122 RFLAYSAQVQF---GYRLGVSVPFLSFRDMLR-AGP--QLLKQAWQSVYQSVRFIE 175
QY 184 DPKLVEILNYFIKVVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELG 243
DB 176 DEHLRQAFSPHSLVGVGNPFTTSIYTLIHALEREWGVPFGGTGALVNGMWKLTDLG 235
QY 244 VEIRLDAEVSIOQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLRS-PASELK-- 300
DB 236 GEIHLNARVEELVADNRVSVRLADGRIFETDVAASNADVNTYKLLGTIPVQGRRA 295
QY 301 KQVFPSCGLVHLGVDRLYPQLAHNPFYSDPHREHDAVFKSHRLSDDDPTIYLVAP 360
DB 296 RLERKSMNSLFLVLYFGLNQPHSLAHTTICFGPRYRELIDEIFTGSALADDFSLYHSP 355
QY 361 CKTDPQAPACCEIKILPHIPHLDPDKLLTAEDYSALRERVVLKLERGLTDLRQHI 420
DB 356 CVTDPSTLAPPCCASFYVLAPVPHLGNAPLQWAGPKLRIDFYDEERYMPGLRSOLVT 415
QY 421 BEYMTPLDIQAKYYSNCGSIY-----GVVADRFRKNLGFAPQRSSELSNLYFVGGSV 472
DB 416 QRIFTQTSRHWAIILGSLFIEPPSLTQGLFA-----ANATRSNLYLVAAGT 464
QY 473 NPGGMPMVTLSGQVLRDKIVADLQ 497
DB 465 HPGAGIPGWGLAESTASLTIEDLQ 489
```

RESULT 8

US-07-783-705A-4

Sequence 4, Application US/07783705A

Patent No. 5429939

GENERAL INFORMATION:

APPLICANT: Misawa, No. 5429939ihiko
APPLICANT: Kobayashi, Kazuo
APPLICANT: Nakamura, Katsumi
APPLICANT: Yamano, Shigeyuki
TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS
NUMBER OF SEQUENCES: 18

b 242 ARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVHSHYRDLGHTRGRTKAAILNR-Q 300
 y 304 RFEPCSGVLHLGVDRLYPQLAHNFFYSDHPRHFDVAFKSHLSDDPTIYLVAPOKT 363
 b 301 RW--SMSFLVHLFGSKRPENLAHSHVIFGPYKGLVNEIFNGPRLPDDESMYLHSPCVT 358
 y 364 DPAQAPAGCEIIKILPHIEL--DPDKLLTAEDYSALRERVVLKLBWGLTDLRQHIVT 420
 b 359 DPLAPEGSTHYVLAPVPHLGRADVWEABAPGYA---ERIFESLERRAIPDLRKHIVT 415
 y 421 EEWYTPLDIQAKYYSNOGSIYGVADRFRNGLFKAPQRSSELSNLYFYGGSVNPGGMPM 480
 b 416 SRIFSPADFSTELSAHSGAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
 y 481 VTLSGGQLVRDKIVADL 496
 b 476 VVGSAKATAQWMLSDL 491

RESULT 10

S-09-298-718-5
 Sequence 5, Application US/09298718
 Patent No. 6124113
 GENERAL INFORMATION:
 APPLICANT: Hohmann, Hans-Peter
 APPLICANT: Pasamontes, Luis
 APPLICANT: Tessier, Michel
 APPLICANT: van Loon, Adolphus
 TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hoffmann-La Roche Inc.
 STREET: 340 Kingsland Street
 CITY: Nutley
 STATE: NJ
 COUNTRY: USA
 ZIP: 07110

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/298,718
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/660,645
 FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.
 REGISTRATION NUMBER: 32,748
 REFERENCE/DOCKET NUMBER: RAN 6002/170
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (201) 235-5801
 TELEFAX: (201) 235-2363
 INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 494 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

S-09-298-718-5

Query Match 24.5%; Score 637; DB 3; Length 494;
 Best Local Similarity 31.2%; Pred. No. 1.1e-56;
 Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;

y 10 IVIGAGLGLSALISATAGFSVQLLEKDKVCGKLNINTKDGFTEDLGSPSLTMTPLHFE 69

b 5 IVIGAGFGGLALAIQLQSAGIATTIVEARDKPGGRAYVWNDQGHVFDAGPTVTDPSLR 64

Qy 70 ALFTGAGKWDYVQIQYEPHNRNFEFGSDVIDLCEDAETORRELDKLGPGTYAQFORP 129
 Db 65 ELWALSGQPMERDVTLLPSPFFYRLTWADGRSFEYVNDDELIQVASFNPADVDGYRFP 124
 Qy 130 LDYSKNLCTETEAGYFAKGLDGFWDLLKFGPLRSLSPDFRSMKDGQVRRFISDPKLYE 189
 Db 125 HDYABEVYRE--GYLKLGTTPPELKLQOMLNAAPALRLQAYRSVHSMVARPIQDPHLRQ 181
 Qy 190 ILNYFKIYVCGSSYPDAPALMNLPLPIQYHYGLVYKGMVGMQAQMEKLAVELGVIRLD 249
 Db 182 AFGFHTLLVCGNPFSTSIYALHALLEREGGVWFAKGGTNQLVAGMVALFERLGGTLLLN 241
 Qy 250 ABYSEIQKQDGRACAVKLANGDVLPAIDIVVSNMEVIPAMEKLL-----RSPASELKQKQ 303
 Db 242 ARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVHSHYRDLGHTRGRTKAAILNR-Q 300
 Qy 304 RFEPCSGVLHLGVDRLYPQLAHNFFYSDHPRHFDVAFKSHLSDDPTIYLVAPOKT 363
 Db 301 RW--SMSFLVHLFGSKRPENLAHSHVIFGPYKGLVNEIFNGPRLPDDESMYLHSPCVT 358
 Qy 364 DPAQAPAGCEIIKILPHIEL--DPDKLLTAEDYSALRERVVLKLBWGLTDLRQHIVT 420
 Db 359 DPLAPEGSTHYVLAPVPHLGRADVWEABAPGYA---ERIFESLERRAIPDLRKHIVT 415
 Qy 421 EEWYTPLDIQAKYYSNOGSIYGVADRFRNGLFKAPQRSSELSNLYFYGGSVNPGGMPM 480
 Db 416 SRIFSPADFSTELSAHSGAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
 Qy 481 VTLSGGQLVRDKIVADL 496
 Db 476 VVGSAKATAQWMLSDL 491

RESULT 11

US-09-546-969-5
 Sequence 5, Application US/09546969
 Patent No. 6207409
 GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/546,969

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 494 amino acids

LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
-09-547-267-5

Query Match 24.5%; Score 637; DB 4; Length 494;

Best Local Similarity 31.2%; Pred. No. 1.1e-56;

Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;

```
10 IVIGAGLGLSAAISLATAGSVQVLIENKDKVGGKLNIMTKDGTTFDLGPSILTMHPHF 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5 IVIGAGFGLALAIRLQASGIATTIVEARDKPGGRAYVWMDQGHVFDAGTIVTDPDSLR 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 ALFTGAGNMADYVQIQKVEPHNFFEDSGSVIDLCEDAETQRRRLDKLGGTVYAOQRF 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 ELWALSQGMERDVTLLPVSPFYRLTWADGRSFEVYVNDDELIRQVASFNPADVDGYRF 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 LDYSKNLCTETAGYFAKGLDGFMDLLKFGPLRSLLSPDVFRRSMQGVRRFISDPKIVE 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 HDYAEVYRE--GYLKLGTTPFLKGLQMLNAAPALMRLQAYRSVHSMVARFIQDPHLRQ 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 ILNFIKTVGSSPDAPALMNLPIYQVHGLVYVKGMYGMAQAMEKLAVELGVEIRLD 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 AFSFHTLLVGNPSTSSIVYALIHAEERGGVWFAKGTQNLVAGVVALPERLGGTLLLN 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 AEVSEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSPASELKMQ 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 ARVTRIDTEGRATGVILLDGRQLRADTVASNGDVHMSYRDLLGHTRRGRTKAAILNR-Q 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 RFEPSCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVAPCKT 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 RW--SMSLFLVHLFGLSKRPELNAHSHSVIPGRYKGLVNEIFNGRPLPDDFSMYLHSPCVT 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 DPAQAPACEIIKILPHIPLH---DPDKLLTAEDYSALRERVLVKLRMGLTDLRQHIVT 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 DPLAPEGMSTHYVLAPVHLGRADVWEAEAPGYA---ERIPLELERRAIPDLRKHILT 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 BEYWTPLDIQKYNSQSGSIYGVVADRFKNGLPKAPORSSELNLYFVGSGVNPQGMFM 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 SRIFSPADFSFELSAHSGSAFSEPIITQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 VTLSQGLVRDKIVADL 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
476 VVGSAKATAQVWLSDL 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14

-09-920-923B-4

Sequence 4, Application US/09920923B

Patent No. 6571134

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Fermentative Carotenoid Production

FILE REFERENCE: 15464 US (C38435/125944)

CURRENT APPLICATION NUMBER: US/09/920,923B

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832

PRIOR FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patent in Version 3.1

SEQ ID NO 4

LENGTH: 494

TYPE: PR

ORGANISM: Flavobacterium sp. R1534

-09-920-923B-4

Query Match

Best Local Similarity 24.5%; Score 637; DB 4; Length 494;

Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;

```
QY 10 IVIGAGLGLSAAISLATAGSVQVLIENKDKVGGKLNIMTKDGTTFDLGPSILTMHPHF 69
Db 5 IVIGAGFGLALAIRLQASGIATTIVEARDKPGGRAYVWMDQGHVFDAGTIVTDPDSLR 64
QY 70 ALFTGAGNMADYVQIQKVEPHNFFEDSGSVIDLCEDAETQRRRLDKLGGTVYAOQRF 129
Db 65 ELWALSQGMERDVTLLPVSPFYRLTWADGRSFEVYVNDDELIRQVASFNPADVDGYRF 124
QY 130 LDYSKNLCTETAGYFAKGLDGFMDLLKFGPLRSLLSPDVFRRSMQGVRRFISDPKIVE 189
Db 125 HDYAEVYRE--GYLKLGTTPFLKGLQMLNAAPALMRLQAYRSVHSMVARFIQDPHLRQ 181
QY 190 ILNFIKTVGSSPDAPALMNLPIYQVHGLVYVKGMYGMAQAMEKLAVELGVEIRLD 249
Db 182 AFSFHTLLVGNPSTSSIVYALIHAEERGGVWFAKGTQNLVAGVVALPERLGGTLLLN 241
QY 250 AEVSEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSPASELKMQ 303
Db 242 ARVTRIDTEGRATGVILLDGRQLRADTVASNGDVHMSYRDLLGHTRRGRTKAAILNR-Q 300
QY 304 RFEPSCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVAPCKT 363
Db 301 RW--SMSLFLVHLFGLSKRPELNAHSHSVIPGRYKGLVNEIFNGRPLPDDFSMYLHSPCVT 358
QY 364 DPAQAPACEIIKILPHIPLH---DPDKLLTAEDYSALRERVLVKLRMGLTDLRQHIVT 420
Db 359 DPLAPEGMSTHYVLAPVHLGRADVWEAEAPGYA---ERIPLELERRAIPDLRKHILT 415
QY 421 BEYWTPLDIQKYNSQSGSIYGVVADRFKNGLPKAPORSSELNLYFVGSGVNPQGMFM 480
Db 416 SRIFSPADFSFELSAHSGSAFSEPIITQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
QY 481 VTLSQGLVRDKIVADL 496
Db 476 VVGSAKATAQVWLSDL 491
```

RESULT 15

US-09-091-725-17

Sequence 17, Application US/09091725

Patent No. 6329141

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Improved methods for transforming Phaffia

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster LLP

STREET: 2000 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: DC

COUNTRY: United States of America

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/091,725

FILING DATE: 23-DEC-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95203620.0

FILING DATE: 22-DEC-1995

APPLICATION NUMBER: EP 96200943.7

FILING DATE: 11-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: E. Victor Donahue

REGISTRATION NUMBER: 35,492

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 582 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
S-09-091-725-17

Query Match 23.8%; Score 618.5; DB 4; Length 582;
Best Local Similarity 28.7%; Pred. No. 1.1e-54; Mismatches 229; Indels 33; Gaps 10;
Matches 148; Conservative 106; Mismatches 229; Indels 33; Gaps 10;
y 5 DNQRVIVIGAGLGLSALATAGFVSQVLIKNDKVGKLNIMTKDGTFTDLGPSILTM 64
b 8 DKPTAIIVGCGIGIATAARLAKGQVTVFEKNDVSGGRCSLIERDGVYRFDQGPSLILL 67
y 65 PHIFPALFTGAGKKNADYVQIQVVEPHNFPEDGSVIDLCEASTORRELDKL-GPCY 123
b 68 PDLEFKOTFEDLGSKMEDVDLIKCEPNYVCHPHDEETFTFSDMALLAKREVERFEKDG 127
y 124 AQQRFLDYSKNLCTETEAGYFAKGLDGFWDLIK--FYGLRSLLSFDVFRSMQGVRRF 181
b 128 DRFLSIQERHRAHVELAVVHVLQKNFPGFAFLRLQFIG--QILALHFFESINTRVCKY 184
y 182 ISDPKLVEILNYFIKIVGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAV- 240
b 185 FKTDRLRRVFSFAVMYMGOSPYAPGTYSLLQYTELGEIWTYRGGFWQVNTLLQIVKR 244
y 241 -ELGVEIRLDAEYSEI--QKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKILRSPA- 296
b 245 NNPSAKFNFNAPVSQVLLSPAKDRATGVRLSEGEHHDVIVNADLVYASEHLIPDDAR 304
y 297 -----SELKKM-----ORFPCSGGLVHLGVDRLYPOLAHNFFYSDHPREHEDA 342
b 305 NKIGQLGEVRRSWADIVGGKKLKGSCSSLSFYWSMDRIVDGLGGHNIFLABDFKGSFDT 364
y 343 VFKSHRLSDPTLYLVAPCKTDPQAQAGCEIILPHIPHLDPDKLLTAEDYSAL---- 398
b 365 IFELGLPADPSFYVNVPSRIDPSAPEGKDAIVILVPCGHIDAS---NPQYNKLIVARA 421
y 399 RERYLVKLE-RMGLTDLRQHIVTEEYWTPLDIOAKYVSNQSGSIYGVVADRFXNLGFKAPQ 457
b 422 RKFVIQTLAKLGLPDPFKMVAEKVHDAPSWKEFNLKDGSTLGLAHNFMQVLGFREPT 481
y 458 RSSELSNLYFVGGSVNEGGMPWTLSCQLVRDKIV 493
b 482 RHPKYDKLFFVGASTHEFTGTGVPVILAGAKULTANQVL 517

Search completed: February 29, 2004, 14:54:53
Job time : 17.1029 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

on on: February 29, 2004, 14:26:38 ; Search time 58.7617 Seconds
(without alignments)
2389.754 Million cell updates/sec

file: US-09-941-947A-24

effect score: 2598
sequence: 1 MNSNDNQNVIVIGAGLGLIS.....MPMTLSQGLVRDKIVADLQ 497

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atadase : A_Geneseqp29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2598	100.0	497	5	Aae222310 Methylo
2	2598	100.0	497	5	Abg61589 High grow
3	2598	100.0	497	5	Aau80333 Methylo
4	2598	100.0	497	6	Ada14538 Methylo
5	1289	49.6	497	6	Ada14544 Staphyloc
6	1289	49.6	500	6	Abm72616 Staphyloc
7	1287	49.5	497	6	Abu16031 Protein e
8	1280	49.3	502	6	Abu43877 Protein e
9	1039	40.0	439	4	Aau34200 Staphyloc
10	713	27.4	434	6	Abm70123 Photorhab
11	703.5	27.1	489	2	Aar13985 Phytoene
12	703.5	27.1	489	2	Aar13984 Phytoene
13	703.5	27.1	489	2	Aaw01123 Phytoene
14	703.5	27.1	489	2	Aaw00174 Phytoene
15	703.5	27.1	489	2	Aaw32475 Erwinia h
16	702.5	27.0	489	2	Aaw01122 Phytoene
17	702.5	27.0	489	2	Aaw00173 Phytoene
18	702.5	27.0	489	2	Aaw32472 Erwinia h
19	687	26.4	492	5	Aae222314 Pantoea s
20	687	26.4	492	6	Aao16021 Pantoea s
21	687	26.4	492	6	Abp96688 Pantoea s
22	672	25.9	502	6	Abm72613 Staphyloc
23	672	25.9	502	6	Ada14542 Staphyloc
24	671	25.8	492	2	Aar07466 Polypepti
25	671	25.8	492	2	Aaw82257 C. utilis

26	671	25.8	492	2	Aaw87889 Protein e
27	671	25.8	492	2	Aaw99099 Erwinia u
28	671	25.8	492	2	Aay26333 Erwinia u
29	653	25.1	499	2	Aar64269 Anabaena
30	640.5	24.7	582	6	Abp97465 Blakeslea
31	637	24.5	494	2	Aaw00871 Flavobact
32	634	24.4	494	2	Aaw59532 Flavobact
33	618.5	23.8	492	2	Aaw22499 Phaffia d
34	607.5	23.4	511	5	Aae22309 Methylo
35	607.5	23.4	511	5	Abg61588 High grow
36	607.5	23.4	511	5	Aau80332 Methylo
37	607.5	23.4	511	6	Ada14534 Methylo
38	596	22.9	290	2	Aaw77702 Crtd proc
39	558	21.5	526	2	Aar95697 Erythroba
40	554	21.3	490	6	Abu32531 Protein e
41	551	21.2	490	5	Abb49224 Listeria
42	522	20.1	543	5	Aao15521 Micrococc
43	495	19.1	544	5	Aao15518 Agromyces
44	475	18.3	548	4	Aab76640 Corynebact
45	475	18.3	548	4	Aab76641 Corynebact

ALIGNMENTS

RESULT 1
AAE222310 standard; protein: 497 AA.
ID AAE222310
XX AC AAE222310;
XX AC

07-AUG-2003 (revised)
25-JUL-2002 (first entry)
XX DT

Methylomonas 16a sp. diapophytoene dehydrogenase (Crtn2) enzyme.
XX DE

Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;
anti-oxidant; steroid; flavour; fragrance; electro-optic application;
aquaculture; enzyme; diapophytoene dehydrogenase; Crtn2.
XX KW

Methylomonas sp.
XX OS

WO200218617-A2.
XX PN

07-MAR-2002.
XX PD

04-SEP-2001; 2001WO-05027490.
XX PF

01-SEP-2000; 2000US-0229858P.
XX PR

01-SEP-2000; 2000US-0229907P.
XX PR

(DUPO) DU PONT DE NEMOURS & CO E I.
XX PA

Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;
Odom JM, Picataggio SK, Rouviere PE;
XX PI

WPI; 2002-351711/38.
XX DR

N-PSDB; AAD35508.
XX DR

Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by
using microorganisms having a nucleic acid molecule encoding enzymes in
the carotenoid biosynthetic pathway and which metabolize single carbon
substrates.
XX XX

Claim 37; Page 131-133; 156pp; English.
XX PS

The invention relates to a method for producing carotenoid compounds. The
method comprises a transformed metabolising host cell, comprising
suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
encoding an enzyme in the carotenoid biosynthetic pathway, under the
control of regulatory sequences, and contacting the host cell with carbon
substrate to produce a carotenoid compound. The method is useful for
producing carotenoid compounds such as anthraxanthin and astaxanthin, by

using microorganism having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolise single carbon substrates. The carotenoids have potent anti-oxidant properties useful in diet, and aquaculture elements. The carotenoids are also useful as intermediates in the synthesis of steroids flavours and fragrances and compounds for potential electro-optic applications. The present sequence is *Methylobacterium* 16a sp. diaphanochrome dehydrogenase (CrtN2) enzyme used in the invention. (Updated on 07-AUG-2003 to correct OS field.)

Sequence 497 AA;
Query Match 100.0%; Score 2598; DB 5; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.5e-241;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2Y 1 MNSNDNRVIVIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLMIMTKDGTDFLGPS 60
2B 1 MNSNDNRVIVIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLMIMTKDGTDFLGPS 60
2Y 61 ILTMPHIFEALFTGAGKNNMADYVQIQVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
2B 61 ILTMPHIFEALFTGAGKNNMADYVQIQVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
2Y 121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFKYGLPLSLLSFDVFRSMDQGVRR 180
2B 121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFKYGLPLSLLSFDVFRSMDQGVRR 180
2Y 181 FISDPKLVIELINYEIKVVGSSPYDAPALMNLPIQVHYGLVYKGMYGQAAMEKLA 240
2B 181 FISDPKLVIELINYEIKVVGSSPYDAPALMNLPIQVHYGLVYKGMYGQAAMEKLA 240
2Y 241 ELGVEIRLDAVSIIQKQGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLSPASELK 300
2B 241 ELGVEIRLDAVSIIQKQGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLSPASELK 300
2Y 301 KMORPEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDPTIYLAP 360
2B 301 KMORPEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDPTIYLAP 360
2Y 361 CKTDPAPACGCEIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHIVT 420
2B 361 CKTDPAPACGCEIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHIVT 420
2Y 421 EYWTPLDIQAKYTSNQSIVGVVADRFKNGLFKAPQRSSELSNLYPVGGSVNPQGGMPM 480
2B 421 EYWTPLDIQAKYTSNQSIVGVVADRFKNGLFKAPQRSSELSNLYPVGGSVNPQGGMPM 480
2Y 481 VTLSCQLVRDKIVADLQ 497
2B 481 VTLSCQLVRDKIVADLQ 497

RESULT 2
ID ABG61589
AC ABG61589 standard; protein; 497 AA.
RC ABG61589;
XX
DT 07-AUG-2003 (revised)
DT 27-AUG-2002 (first entry)
XX
DE High growth methanotrophic bacterial strain polypeptide #39.
XX
KW High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme; methane; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;
KW pyrophosphate dependent phosphofructokinase; nitrogen-containing compound; ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;
KW methane-containing environment; waste water treatment system; isoprenoid; nitrous oxide; terpenoid; animal feed; exopolysaccharide.
XX
CS *Methylobacterium* sp.
XX
PN WO200220728-A2.

XX 14-MAR-2002.
XX
XX 28-AUG-2001; 2001WO-US026827.
XX
XX 01-SEP-2000; 2000US-0229858P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Koffas M, Odom JM, Schenzle A;
XX
XX WPI; 2002-454358/48.
XX N-PSDB; ABK83268.
XX
XX New high growth methanotrophic bacterial strain, useful for producing single cell proteins, grows on a C1 carbon substrate, and comprises a functional gene encoding in Embden-Meyerhof carbon pathway.
XX
XX Claim 11; Page 153-154; 157pp; English.
XX
XX The invention relates to a high growth methanotrophic bacterial strain, which grows on a C1 carbon substrate e.g. methane and methanol, and comprises a functional Embden-Meyerhof carbon flux pathway comprising a gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S RNA. The bacterial strain is useful for the production of single cell protein and for the biotransformation of a nitrogen-containing compound, e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the production of a feed product comprising a protein, carbohydrates and a pigment and for reducing oxygen demand, for removing nitrates and nitrites in methane-containing environments such as landfills, waste water treatment systems or anywhere the invention can be used as a present. The bacterial strain of the invention can be used as a denitrifying agent for the conversion of nitrate or nitrite to nitrous oxide with methane or methanol as a carbon source. It is also used in the production of biomass including proteins, carbohydrates and a wide variety of pigments (particularly for isoprenoid pigments for the purpose of generating animal feeds), in production of terpenoid and carotenoid compounds, useful as pigments and as monomers in polymeric materials and in production of exopolysaccharides at high levels. Sequences ABG61551-ABG61590 represent high growth methanotrophic bacterial strain proteins of the invention. (Updated on 07-AUG-2003 to correct OS field.)
XX
XX Sequence 497 AA;
SQ

Query Match 100.0%; Score 2598; DB 5; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.5e-241;
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSNDNRVIVIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLMIMTKDGTDFLGPS 60
DB 1 MNSNDNRVIVIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLMIMTKDGTDFLGPS 60
QY 61 ILTMPHIFEALFTGAGKNNMADYVQIQVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
DB 61 ILTMPHIFEALFTGAGKNNMADYVQIQVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
QY 121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFKYGLPLSLLSFDVFRSMDQGVRR 180
DB 121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFKYGLPLSLLSFDVFRSMDQGVRR 180
QY 181 FISDPKLVIELINYEIKVVGSSPYDAPALMNLPIQVHYGLVYKGMYGQAAMEKLA 240
DB 181 FISDPKLVIELINYEIKVVGSSPYDAPALMNLPIQVHYGLVYKGMYGQAAMEKLA 240
QY 241 ELGVEIRLDAVSIIQKQGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLSPASELK 300
DB 241 ELGVEIRLDAVSIIQKQGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLSPASELK 300
QY 301 KMORPEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDPTIYLAP 360
DB 301 KMORPEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDPTIYLAP 360
QY 361 CKTDPAPACGCEIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHIVT 420

361 CKTDPAPACCEIIKILPHILDPKLLTAEDYSALRERVLKLERMGLTLRQHIVT 420
 421 BEYWTPLDIOAKYYSNQSIIYGVVADRFKNGFKAPQSRSELNLYFVGSGVNPGGMPM 480
 421 BEYWTPLDIOAKYYSNQSIIYGVVADRFKNGFKAPQSRSELNLYFVGSGVNPGGMPM 480
 481 VTLSGQLVRDKIVADLQ 497
 481 VTLSGQLVRDKIVADLQ 497

RESULT 3
 AU80333
 AAU80333 standard; protein; 497 AA.

AAU80333;

15-JUL-2002 (first entry)

Methylomonas 16a ORF9 crtN2 protein sequence.

Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;
 keratinoid; pigment; flavour; fragrance; open reading frame 9; ORF9;
 crtN2; crtN copy2; diapophytoene dehydrogenase enzyme.

Methylomonas sp.

WO2000220733-A2.

14-MAR-2002.

29-AUG-2001; 2001WO-US026852.

01-SEP-2000; 2000US-0229907P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Cheng Q, Koffas M, Norton KC, Odom JM, Picataggio SK;
 Rouviere PE, Schenale A, Tomb J;

WPI; 2002-383051/41.
 N-PSDB; ABK50089.

Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
 isolated from Methylomonas 16a, useful for the production of isoprenoid
 compounds.

Claim 4; Page 81-82; 84pp; English.

The present invention relates to a new nucleic acid molecule encoding an
 isoprenoid biosynthetic enzyme isolated from Methylomonas 16a. The
 invention is useful for obtaining a nucleic acid molecule encoding an
 isoprenoid compound biosynthetic enzyme, and for the microbial production
 of isoprenoid compounds. The molecules of the invention are also useful
 for regulating isoprenoid biosynthesis in an organism and for producing
 recombinant organisms for producing various isoprenoid compounds. The
 nucleic acid is also useful for feed additive, for the production of
 keratinoids and their derivatives, isoprenoid intermediates, and as pure
 products useful as pigments, flavours and fragrances. The present amino
 acid sequence represents the Methylomonas 16a open reading frame 9 (ORF9)
 crtN2 (crtN copy2) copy 2 of diapophytoene dehydrogenase enzyme) protein
 of the invention, as described above

Sequence 497 AA;

Query Match 100.0%; Score 2598; DB 5; Length 497;
 Best Local Similarity 100.0%; Pred. No. 1.5e-241;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSNDNQRVIVIGAGLGLSLAISLATAGFSVQLIERKNDKVGKGLNTMTKDGFTDLGSPS 60
 1 MNSNDNQRVIVIGAGLGLSLAISLATAGFSVQLIERKNDKVGKGLNTMTKDGFTDLGSPS 60

QY 61 ILTWPHIPEALFTGAGKNMADYVOIQKVEPHRNFFEDGSGVIDLCEAETORRELDKLG 120
 DB 61 ILTWPHIPEALFTGAGKNMADYVOIQKVEPHRNFFEDGSGVIDLCEAETORRELDKLG 120
 QY 121 GTYAQFQRFLLDYSKNLCTETETAGYFAKGLDGFNDLLKPYGLPSLLSFDVFRSMDQGVRR 180
 DB 121 GTYAQFQRFLLDYSKNLCTETETAGYFAKGLDGFNDLLKPYGLPSLLSFDVFRSMDQGVRR 180
 QY 181 FISDPKLVETILNYPFKYVGSSPYDAPALMNLPIQYHYGLWYVKGGMGMAQAMEKLAV 240
 DB 181 FISDPKLVETILNYPFKYVGSSPYDAPALMNLPIQYHYGLWYVKGGMGMAQAMEKLAV 240
 QY 241 ELGVEIRLDAEVSBIQKQDGRACAVKLANGDVLFPADIVWSNMEVIPAMEKLLRSPASELK 300
 DB 241 ELGVEIRLDAEVSBIQKQDGRACAVKLANGDVLFPADIVWSNMEVIPAMEKLLRSPASELK 300
 QY 301 KMQRPSPSCGLVHLGVDRILYPOLAHNPFYSDHPREHFDVAFKSHRLSDDPTIYLVAP 360
 DB 301 KMQRPSPSCGLVHLGVDRILYPOLAHNPFYSDHPREHFDVAFKSHRLSDDPTIYLVAP 360
 QY 361 CKTDPAPACCEIIKILPHILDPKLLTAEDYSALRERVLKLERMGLTLRQHIVT 420
 DB 361 CKTDPAPACCEIIKILPHILDPKLLTAEDYSALRERVLKLERMGLTLRQHIVT 420
 QY 421 BEYWTPLDIOAKYYSNQSIIYGVVADRFKNGFKAPQSRSELNLYFVGSGVNPGGMPM 480
 DB 421 BEYWTPLDIOAKYYSNQSIIYGVVADRFKNGFKAPQSRSELNLYFVGSGVNPGGMPM 480
 QY 481 VTLSGQLVRDKIVADLQ 497
 DB 481 VTLSGQLVRDKIVADLQ 497

RESULT 4

ADA14538

ID ADA14538 standard; protein; 497 AA.

XX ADA14538;

XX 06-NOV-2003 (first entry)

XX Methylomonas sp. 16a CrtN2 protein SEQ ID NO:8.

XX carotenoid biosynthetic enzyme; CBE;

XX omega-aldehyde-functionalised carotenoid;

XX carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;

XX CrtN2.

XX Methylomonas sp.

XX WO2003068917-A2.

XX 21-AUG-2003.

XX 11-FEB-2003; 2003WO-US004150.

XX 11-FEB-2002; 2002US-0355939P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Cheng Q, Norton KC, Tao L;

XX WPI; 2003-697524/66.

XX N-PSDB; ADA14537.

XX New nucleic acid encoding carotenoid biosynthesis enzymes, useful for
 preparing functionalized carotenoids, e.g. pigments and antioxidants,
 also encoded polypeptides.
 XX Claim 19; Page 99-102; 125pp; English.

The present invention describes nucleic acid sequences encoding carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE polypeptides; (2) a chimeric gene (CG) containing a CBE linked to regulatory sequences; (3) host cell transformed with CG; (4) obtaining a nucleic acid that encodes a CBE; (5) products of method (4); (6) producing an omega-aldehyde-functionalised carotenoid (X); (7) producing an omega-carboxy-functionalised carotenoid (Y); (8) producing C30 carotenoids (Z); and (9) regulating carotenoid biosynthesis in an organism. CBE sequences have antioxidant activity. CBEs can be used for microbial production of aldehyde- or carboxy-functionalised carotenoids and C30 carotenoids, and are potentially useful as pigments and antioxidants, e.g. as feed additives. The present sequence represents a CrtN2 CBE protein isolated from *Methylobacterium* sp. strain 16a, which is used in the exemplification of the present invention.

Query Match 100.0%; Score 2598; DB 6; Length 497;

Best Local Similarity 100.0%; Pred. No. 1.5e-241;

Mismatches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSNDNRVIVIGAGLGLSAAISLATAGSVOLIEKNKVGKLNIMTKDGTEDLGP 60

1 MNSNDNRVIVIGAGLGLSAAISLATAGSVOLIEKNKVGKLNIMTKDGTEDLGP 60

61 ILTWPHFEALFTGAGNMDYVQIOKVEPHNRPFDGSDVLDCEDAETORRELDKLP 120

61 ILTWPHFEALFTGAGNMDYVQIOKVEPHNRPFDGSDVLDCEDAETORRELDKLP 120

121 GTVAQFORFLDYKNLCTETAGYFAKGLDGFWDLLKPYGPRSLSPDFVFRSMQGVRR 180

121 GTVAQFORFLDYKNLCTETAGYFAKGLDGFWDLLKPYGPRSLSPDFVFRSMQGVRR 180

181 FISDPKLVEILNPKYKVGSSPDAPALNLLPYIOHYGLWTVKGMQMAQMEKLA 240

181 FISDPKLVEILNPKYKVGSSPDAPALNLLPYIOHYGLWTVKGMQMAQMEKLA 240

241 ELGVEIRLDAEVSIIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSPASELK 300

241 ELGVEIRLDAEVSIIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSPASELK 300

301 KMQFEPSCSLVHLGVDRILYPOLAHNPFYSDHPREHDAVFKSHRSLSDPTIYLVP 360

301 KMQFEPSCSLVHLGVDRILYPOLAHNPFYSDHPREHDAVFKSHRSLSDPTIYLVP 360

361 CKTDPAQAPACCEIIKILPHIPLDPPDKLLTAEDYSALRERVLVGLRMGLTDLRQHI 420

361 CKTDPAQAPACCEIIKILPHIPLDPPDKLLTAEDYSALRERVLVGLRMGLTDLRQHI 420

421 BEYWTPLDIOAKYYSNQGSIYGVVADRFKLGPKAPQSSSELSNLYFVGSGVNPGGMPM 480

421 BEYWTPLDIOAKYYSNQGSIYGVVADRFKLGPKAPQSSSELSNLYFVGSGVNPGGMPM 480

481 VTLGSQLVRDKIVADLQ 497

481 VTLGSQLVRDKIVADLQ 497

RESULT 5

ID ADAL4544

ADAL4544 standard; protein; 497 AA.

ADAL4544;

06-NOV-2003 (first entry)

Staphylococcus aureus CrtN2 protein SEQ ID NO:14.

carotenoid biosynthetic enzyme; CBE;

omega-aldehyde-functionalised carotenoid;

omega-carboxy-functionalised carotenoid; C30 carotenoid;

carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;

CrtN2.

XX Staphylococcus aureus.
OS WC2003068917-A2.
FN 21-AUG-2003.
PD 11-FEB-2003; 2003WO-US004150.
XX 11-FEB-2002; 2002US-0355939P.
PR (DUPO) DU PONT DE NEMOURS & CO E I.
XX Cheng Q, Norton KC, Tao L;
PI WPI; 2003-697524/66.
XX N-PSDB; ADA14543.
DR New nucleic acid encoding carotenoid biosynthesis enzymes, useful for
XX preparing functionalized carotenoids, e.g. pigments and antioxidants,
XX also encoded polypeptides.
XX Claim 19; Page 114-117; 125pp; English.
XX The present invention describes nucleic acid sequences encoding
XX carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE
XX polypeptides; (2) a chimeric gene (CG) containing a CBE linked to
XX regulatory sequences; (3) host cell transformed with CG; (4) obtaining a
XX nucleic acid that encodes a CBE; (5) products of method (4); (6)
XX producing an omega-aldehyde-functionalised carotenoid (X); (7) producing
XX an omega-carboxy-functionalised carotenoid (Y); (8) producing C30
XX carotenoids (Z); and (9) regulating carotenoid biosynthesis in an
XX organism. CBE sequences have antioxidant activity. CBEs can be used for
XX microbial production of aldehyde- or carboxy-functionalised carotenoids
XX and C30 carotenoids, and are potentially useful as pigments and
XX antioxidants, e.g. as feed additives. The present sequence represents a
XX CrtN2 CBE protein isolated from *Staphylococcus aureus*, which is used in
XX the exemplification of the present invention.
XX Sequence 497 AA:
Query Match 49.6%; Score 1289; DB 6; Length 497;
Best Local Similarity 51.9%; Pred. No. 3.7e-115;
Matches 254; Conservative 85; Mismatches 148; Indels 2; Gaps 2;
QY 7 QRVIVIGAGLGLSAAISLATAGSVOLIEKNKVGKLNIMTKDGTEDLGP 66
Db 3 KHIIVIGAGLGLSAAISLATAGSVOLIEKNKVGKLNIMTKDGTEDLGP 62
QY 67 IFEALFTGAGNMDYVQIOKVEPHNRPFDGSDVLDCEDAETORRELDKLP 126
Db 63 IFEALFTGAGNMDYVQIOKVEPHNRPFDGSDVLDCEDAETORRELDKLP 122
QY 127 QRFIDYKNLCTETAGYFAKGLDGFWDLLKPYGPRSLSPDFVFRSMQGVRR 186
Db 123 QNVLNTRRIDRITEKGYFNYGLDLSQIIKFHGPLNALLINYDVHTMQQAIDKRISN 182
QY 187 LVEILNPKYKVGSSPDAPALNLLPYIOHYGLWTVKGMQMAQMEKLA 246
Db 183 LRQMLGTFIKYKVGSSPDAPALNLLPYIOHYGLWTVKGMQMAQMEKLA 242
QY 247 RLDAEVSIIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSPASELK 305
Db 243 HTGARVDNPKYKVGSSPDAPALNLLPYIOHYGLWTVKGMQMAQMEKLA 302
QY 306 EPSCSGLVHLGVDRILYPOLAHNPFYSDHPREHDAVFKSHRSLSDPTIYL 365
Db 303 EPASSGYVHLGVDRILYPOLAHNPFYSDHPREHDAVFKSHRSLSDPTIYL 362
QY 366 AQAPACCEIIKILPHIPLDPPDKLLTAEDYSALRERVLVGLRMGLTDLRQHI 425
Db 363 TCAPGVENIKVLPHPVQ-DQPTTEDYAKPRDKILDKLKMGLTDLRKHIIYEDVMT 421

426 PLDIQAKYNSQSIYGVVADRFKXGLGFKAPQRSSELSNLYFVGGSNPQGGPMVTLTG 485
422 PEDIEKTRSRNGAIYGVVADKXKNGKFKPKESQYFENLYFVGGSNPQGGPMVTLTG 481
486 QLVDRKIVA 494
482 QQVADKINA 490
RESULT 6
ABM72616 standard; protein; 500 AA.
ABM72616;
20-NOV-2003 (first entry)
Staphylococcus aureus protein #1856.
Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
enzymatic assay; antibiotic target.
Staphylococcus aureus.
WO200294863-A2.
28-NOV-2002.
27-MAR-2002; 2002WO-IB002637.
27-MAR-2001; 2001GB-00007661.
(CHIR-) CHIRON SPA.
Masignani V, Mora M, Scarselli M;
WPI: 2003-120786/11.
N-PSDB; AC74176.
New Staphylococcus aureus protein, useful as a vaccine for treating or
preventing Staphylococcal infection, specifically an infection caused by
S. aureus, e.g. sepsis.
Claim 1; SEQ ID NO 3712; 49pp; English.
The invention relates to novel genes and encoded proteins from
Staphylococcus aureus. A composition comprising the S. aureus protein, a
nucleic acid encoding the protein, or an antibody to the protein, is
useful as a pharmaceutical, particularly as a vaccine for treating or
preventing infection due to Staphylococcus bacteria, specifically an
infection caused by S. aureus. The composition is particularly useful for
treating or preventing sepsis in a patient. The composition can also be
used for diagnostics. The protein is also used in an assay for enzymatic
studies and as a target for antibiotics. This sequence represents one of
the novel S. aureus proteins of the invention
Sequence 500 AA;
Query Match 49.6%; Score 1289; DB 6; Length 500;
Best Local Similarity 51.9%; Pred. No. 3.8e-115;
Matches 254; Conservative 85; Mismatches 148; Indels 2; Gaps 2;
7 QRIVVIGAGLGLSAAISLATAGFSVQLIERNDKVGKLNMTXDTGTFDLGPSILTMPH 66
6 KHIIVIGLGLGISAIRMAQSGYSVLYEQNNHIGKVNHRHESDGFGLGPSILTMPY 65
67 IPEALFTGAGKMDYVQIQKVEHWRNFFRDSGVLDICDAETQRELDKLGCTVQAF 126
66 IFEKLFYSKKQMSDYVTKRLPHQWRSFDPDGTIDLYEGIKETGQHNAILSKQDIEEL 125
127 QRLDYSKMLCTETAGYFAGKLDGFDLLKPYGLRSLLSFDFRSNDQGVRRFISDPK 186
126 QYLYNTRIDRITEKGFVNGLOTLISQIIRKHPFLNALINYDVHTMQAIDKRLSNPY 185

187 LVEILNLYPIKTVGSSPYDAPALMNLPIYQYEGLYWYVKGGMGMAQAMEKLAVELGVEI 246
186 LRQMLGYPIKTVGSSSYDAPAVLSNLFHMQOQGLWYVGGIHHLANALEKLAREEGVTI 245
247 RLDAEVSIOKODGRACAVKLANGVLPADIVSNMEVI PAMEKILRS PASELKMQR-F 305
246 HTGARVDNIKTYYQRRVTGVRDGTGEFVKADYIIISNMEVIPTKYLIHLDTQRLNKLREF 305
306 EPSCSLVHLGVDRLYPOLAHNFFYSDHPREHFDAVFKSHRSLSDPTIYLVAPCKTDP 365
306 EPASGYVHLGVACQYPOLAHNFFFTENAYLNYQQVFEKVLDPDPTIYLVVNTKTDH 365
366 AQAPAGCEIHKILPHIPHLDPDKLLTAEDYSALSERVLKLERMGLTDLROHIVTEEYWT 425
366 TQAPVGENIKVLPHIPYIQ-DQPTTETDYAKFRDKILDKLEKMGTLDLRKHIIYEDVWT 424
426 PEDIQAKYNSQSIYGVVADRFKXGLGFKAPQRSSELSNLYFVGGSNPQGGPMVTLTG 485
425 PEDIEKTRSRNGAIYGVVADKXKNGKFKPKESQYFENLYFVGGSNPQGGPMVTLTG 484
486 QLVDRKIVA 494
485 QQVADKINA 493
RESULT 7
ABU16031
ID ABU16031 standard; protein; 497 AA.
AC ABU16031;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #1558.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200277183-A2.
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI: 2003-029926/02.
DR N-PSDB; ACAL19901.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 43955; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX Photorhabdus luminescens.
 OS WO200234867-A2.
 PN 28-NOV-2002.
 XX 07-FEB-2002; 2002WO-IB003040.
 XX 07-FEB-2001; 2001FR-00001659.
 XX (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchrieser C;
 XX WPI; 2003-148459/14.
 DR Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX Claim 2; SEQ ID NO 3220; 1205pp; French.
 XX The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX Sequence 494 AA;
 SQ

Query Match
 Best Local Similarity 27.4%; Score 713; DB 6; Length 494;
 Matches 158; Conservative 97; Mismatches 231; Indels 6; Gaps 3;
 8 RVIVTACGLGSLAAISATAGFSVQLERKNDKVGKLNINWKDGFDFDGLPSILTWPHI 67
 3 KALVICAGFGGLALALQSGAIPICILEQRDKPGRAYVYKEQGFEDAGFTVITAPNV 62
 68 FEALFTGAKMADYVOIKVEPHWNPFDGSSVLDLCEADETORRELDKLGPGTYAQFO 127
 63 IELFTQAGKGMADYVDLPPVHPFFRLCWESGKMPFDNDQOHLAQHTNFDVNGYR 122
 128 RFLDYSKNLCTETEAGYFAKGLDGFMDLKFYGLRSLLSDFVFRSMDQGVRRPISPKL 187
 123 RFLDYSRAFNK---GYLKLGTVPFLSPDNLAAQPLIRLEAKRWVSQVARRIKDES 179
 188 VEILNFFIKYGVSSYDAPALMNLPLFYQHYGLWYKGMGKQAQAMEKLAVELGVIR 247
 180 RQAFSHSLIGGNPFAASSYITLHALERWGWFPFRGGTSALVVEAWKLFDTIGGEIE 239

QY 248 LDAEVSIEIQKQDGRACAVKLANGDVLDPADIVVSNNEVIPAMEKLL-RSPA--SELKQOR 304
 DB 240 LNAKVKRETTNGNRTVTGQLADGNMTCDVVASNADVIHTYKLLGQHPVGIAARTLVR 299
 QY 305 FEPSCSGVLHLGVDRLYPOLAHENPFYSDHPREHFDVAFKSHRLSDDPITYLVAPCKTD 364
 DB 300 KRMNSLFLVYFGLNHHHTQLAHTVTCFGRPKYELIEDIFVHDLSEDFSLYLHAPSVTD 359
 QY 365 PAQAPAGCEIIKILPHIPHLDPDKLLTAEDYSALRERVVLKLERMGLTDLROHIVTEYW 424
 DB 360 PSLAPRGCAVYVTLAPVPHLGTANLWDIEGPRLDRIFFAYLEKYIMPGLLKQVWHRIF 419
 QY 425 TPLDIQAKYVNSQSGIYGVVADRFKNLGFKAPOSSSLSNLYFVGGSVNPGGMPVWTL 484
 DB 420 TFFDFRDQINHLGSAFSFEPFLITQSAWFRPHNRDNRLNLYLVGAGTHFGAGIPGVIGS 479
 QY 485 GOLVRDKIVADL 496
 DB 480 AKATATLMLEDI 491
 RESULT 11
 AARI3985
 ID AARI3985 standard; protein; 489 AA.
 XX AC
 XX AARI3985;
 DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-NOV-1991 (first entry)
 XX Phytoene_dehydrogenase-4H from pARC146D.
 XX GGPP; carotenoid; phytoene; zeaxanthin; lycopene.
 XX Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39368).
 XX WC9113078-A.
 XX 05-SEP-1991.
 XX 04-MAR-1991; 91US-US0001458.
 XX 02-MAR-1990; 90US-00487613.
 XX 18-MAY-1990; 90US-00525551.
 XX 03-AUG-1990; 90US-00562674.
 XX 28-FEB-1991; 91US-00662921.
 XX (STAD) AMOCO CORP.
 XX Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;
 PI Yen HC;
 XX WPI; 1991-281410/38.
 XX N-PSDB; AAQ13720.
 XX Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA
 PT encoding enzymes from Erwinia herbicola.
 XX Disclosure; Fig 15(1-4); 313pp; English.
 XX There are a total of six relevant genes in a 7900 bp region that cause E.
 CC coli cells to produce Gppp and the carotenoids phytoene through
 CC zeaxanthin diglucoside, which is the final prod. identified in the
 CC carotenoid pathway contd. in plasmid pARC376 (contg. a ca. 13 kb
 CC chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607
 CC (1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase, beta-
 CC phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-
 CC carotene hydroxylase, and zeaxanthin glycosylase are represented in
 CC AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726
 CC respectively. Recombinant expression plasmids can be used to produce
 CC large ants. of the enzymes and hence large amt. of the carotenoids which
 CC they synthesise. Carotenoids are pigments with a variety of applications.

AAW01123 standard; protein; 489 AA.
AAW01123;
16-OCT-2003 (revised)
25-MAR-2003 (revised)
11-DEC-1996 (first entry)
Phytoene dehydrogenase-4H encoded on pARC146D.
GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
pigment; food colourant; chloroplast transit peptide; increase yield;
tobacco ribulose bis-phosphate carboxylase-oxygenase.
Pantoea agglomerans.
US5530188-A.
25-JUN-1996.
21-JUL-1993; 93US-00095726.
02-MAR-1990; 90US-00487613.
18-MAY-1990; 90US-00525551.
03-AUG-1990; 90US-00562674.
28-FEB-1991; 91US-00662921.
30-OCT-1991; 91US-00785566.
(STAD) AMOCO CORP.
Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
Ausich RL;
WPI; 1996-308823/31.
N-PSDB; AAT40793.
Increasing prodn. of total carotenoid(s) in a higher plant - by
transforming with vector encoding chloroplast transit peptide operably
linked to the Erwinia herbicola lycopene cyclase structural gene.
Example 12; Col 97-100; 93pp; English.
The present sequence is that of a recombinant Erwinia herbicola phytoene
dehydrogenase-4H (from pARC146D) which produces lycopene biosynthetically
from phytoene through four sequential dehydrogenation reactions. Other
enzymes involved in the carotenoid biosynthesis pathway include
geranylgeranyl pyrophosphate (AAW01119), phytoene synthase (W01121) and
lycopene cyclase (AAW01125). Manipulation of, in part., lycopene
cyclase gene (AAW01124) of the tobacco ribulose bis-phosphate carboxylase-
peptide (AAW01124) can lead to increased production of total carotenoids in
the chloroplast of transformed plants as compared to native, non-
transformed plants of the same type. Beta-carotene is an effective and
apparently harmless food colourant and is also in the pathway for
biological synthesis of further C40 carotenoids such as zeaxanthin and
zeaxanthin diglucoside. (Updated on 25-MAR-2003 to correct PF field.)
(Updated on 16-OCT-2003 to standardise OS field)
Sequence 489 AA;
Query Match 27.1%; Score 703.5; DE 2; Length 489;
Best Local Similarity 32.9%; Pred. No. 1.3e-58;
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;
7 QRVIVIGAGLGSNAISLATAGSVOLIKNDKVGKXIMTKGFTFDLGSILTMPEH 66
2 EKTWVIGAGLGSNAISLATAGSVOLIKNDKVGKXIMTKGFTFDLGSILTMPEH 61
67 IFEALFTGAGKNADYVQIKVEPHWRNFEDGSDVIDCEDAEFQRRELKLGPGTYAQF 126
62 ALREALFTLAGRMEDYVRLPVKFFRLCWESGKTLVDYANDSPLEAQITQFNPRDVEG 121

QY 127 QRFLDYSKNLCTETETAGYFAKG---LDGEWDLLKFKYGLRSLLSFDVFRSMQGVRRFIS 183
DB 122 RRFAYSQAVFQE---GYLRLGSPVFLSPRDLR-AGP---QLLKLQAWQSVTQSVSRFIE 175
QY 184 DPKLVEILNTFIKYVGSSPYDAPALNNLLPYIQHYCLAVYKGYGMAQAKELAVELG 243
DB 176 DEHLRQAFSEHSLLVGGNPFTTSSITLIALERBWGVWPFPEGGTGLVNGWVKLFTDLG 235
QY 244 VEIRLDAVSEIKQKQGRACAVKLANGDVLPAIVVSNMEVTPAKELKLS-PASBLK-- 300
DB 236 GEIELNARVELLVADNRVSQVELADGRIPDTDAVASNADVNVTKKLLGTIPVGKRAA 295
QY 301 KMQRPEPSCGLVHLGVDRLYQLAHENFFYSDHPREHFDVFKSHRSLSDOPTIYLVAP 360
DB 296 RLEKSMNSLFLVLTGLNQPHSQLAHTICFQPYRELIIDEFTGSALADDFSLYHSP 355
QY 361 CKTDPAQAPAGCEIILPHIPHLDPDKLITADYSGALZERVLVKLERMGLTDLRQIYV 420
DB 356 CVTDPSLAPPFCASFVYLAAPVHLGNAPLDWAQSGPKLRDRIPDYLEERYMPCGLRSQVLT 415
QY 421 EYWTPLDIQAKYVSNQSIY-----GVVADRFKNLGFKAQPSSELSNLYFVGGSV 472
DB 416 QRIFTQTSRHWIAIILGSLFIEPSPILTQGLFA-----ANATRHNSNLYLVAAGT 464
QY 473 NPGGEMPMVTLGQLVRDKIVADIQ 497
DB 465 HPGAGIFGVVGLAESTRASLMIEDLQ 489
RESULT 14
AAW00174
ID AAW00174 standard; protein; 489 AA.
XX AC AAW00174;
XX DT 16-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 17-OCT-1996 (first entry)
XX DE Phytoene dehydrogenase-4H encoded by DNA derived from pARC146D.
XX KW E. herbicola; geranylgeranyl pyrophosphate synthase; pARC376;
KW GGPP synthase; biosynthesis; carotenoid; lycopene;
KW farnesyl pyrophosphate; phytoene; FPP; isopentenyl pyrophosphate; IPP;
KW tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;
XX KW food colourant; herbicide; norflurazon.
XX OS Pantoea agglomerans.
XX PN US5530189-A.
XX PD 25-JUN-1996.
XX PF 22-JUL-1993; 93US-00096043.
XX PR 02-MAR-1990; 90US-00487613.
XX PR 18-MAY-1990; 90US-00525551.
XX PR 03-AUG-1990; 90US-00562674.
XX PR 28-FEB-1991; 91US-00662921.
XX PR 30-OCT-1991; 91US-00785568.
XX (STAD) AMOCO CORP.
XX PA Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
XX PI Ausich RL;
XX DR WPI; 1996-308824/31.
XX DR N-PSDB; AAT37095.
XX PT DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for
XX PT prodn. of lycopene, and to produce transgenic plants resistant to
XX PT norflurazon.

296 RLERKXNSNLFVLYFGLNQPHSOLAHTICGPRYRELIDEIFTGSALADDFSLYHSP 355
361 CKTDPQAQAPACCEIIKILPHIPELDPDKLLTAEVDSALRERVVLKLERMGLTDLRQHIVT 420
356 CVTDPSLAPPCASFFVLAAPVPHLGNAPLDWAGEGPKLRDRIFDYLEERYMPLRSSQLVT 415
421 ERYWTLPLDIQAKYYSNOGSIIY-----GVVADRFNGLGFKAPQRSSELSNLYFVGGSV 472
416 QRIFTRQTSRRHAWIAIIGSLFIEPPSLTQGLEFA-----ANATRHSNLYLVAAGT 464
473 NPGGGMPEVTLGQLVRDKIVADLQ 497
465 HPGAGIFGVVGLAESTASLMIEDLQ 489

Search completed: February 29, 2004, 14:43:47
Job time : 60.8617 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:33:49 ; Search time 42.0179 Seconds
(without alignments)
3837.172 Million cell updates/sec

cle: US-09-941-947a-22

fect score: 2/11
quence: 1 MANTKHIIIVAGPGGLGAG.....KRVRFKDIASHAWLKAKA 511

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

SPTREMBL 25:*

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phase:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp virus:*
- 16: sp bacteriopl:*
- 17: sp archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1009	37.2	537	16 Q7UF87	Q7uf87 rhodopirell
2	929.5	34.3	490	16 Q8Y8A8	Q8y8a8 listeria mo
3	915.5	33.8	488	16 Q92D21	Q92d21 listeria in
4	883.5	32.6	517	2 QZG34	Q3zge4 heliobacill
5	808	29.8	498	16 Q88W3	Q88w3 lactobacill
6	794.5	29.3	499	16 Q8YKN8	Q8ykn8 anabaena sp
7	783.5	28.9	499	2 Q44318	Q44318 anabaena sp
8	739.5	27.3	511	16 Q8ENM0	Q8enm0 oceanobacil
9	729	26.9	514	2 Q8KZ28	Q8kz28 uncultured
10	726	26.8	513	2 Q840T3	Q840t3 rhodocyclu
11	725	26.7	511	2 Q3UP98	Q3up98 rhodocyclu
12	723	26.7	506	2 Q3UKX2	Q3ukx2 bradyrhizob
13	713.5	26.3	492	2 Q47845	Q47845 pantoea agg
14	705.5	26.0	504	2 Q93C17	Q93c17 xanthobacte
15	703.5	25.9	492	2 Q8GCS1	Q8gcs1 pantoea ste
16	702.5	25.9	502	16 Q99R76	Q99r76 staphylococ

17	702.5	25.9	502	16 Q8NUQ6	Q8nuq6 staphylococ
18	699.5	25.8	506	2 Q07855	Q07855 staphylococ
19	683.5	25.2	492	2 Q8VUJ5	Q8vuj5 pantoea agg
20	673	24.8	582	3 Q13506	Q13506 phaffia rho
21	668	24.6	582	3 Q7Z858	Q7z858 phaffia rho
22	651.5	24.0	548	16 Q9RW08	Q9rw08 deinococcus
23	650	24.0	579	3 Q8X020	Q8x020 gibberella
24	640.5	23.6	494	2 P94790	P94790 flavobacter
25	631	23.3	499	16 Q8XK44	Q8xk44 clostridium
26	628	23.2	579	3 Q9Y7H8	Q9y7h8 mucor circi
27	624.5	23.0	526	2 Q06757	Q06757 erythroba
28	610.5	22.5	501	2 Q9RLH4	Q9rlh4 paracoccus
29	596	22.0	514	17 Q27835	Q27835 methanobac
30	567	20.9	534	16 Q7UJE6	Q7uje6 rhodopirell
31	564.5	20.8	523	16 Q9RX9	Q9rx9 streptomyc
32	552	20.4	499	2 Q9KS66	Q9ks66 mycobacteri
33	547.5	20.2	512	17 Q9HPD8	Q9hpd8 halobacteri
34	540	19.9	501	3 Q7Z857	Q7z857 phaffia rho
35	531	19.6	588	2 Q83X01	Q83x01 streptomyc
36	521.5	19.2	488	2 Q7WT72	Q7wt72 marine bact
37	521	19.2	513	16 Q93HP2	Q93hp2 streptomyc
38	514	19.0	494	16 Q8ENM2	Q8enm2 oceanobacil
39	504	18.6	485	2 Q50893	Q50893 myxococcus
40	502.5	18.5	497	16 Q99R73	Q99r73 staphylococ
41	499.5	18.4	497	16 Q8NUQ3	Q8nuq3 staphylococ
42	485.5	17.9	536	17 Q9HP88	Q9hp88 halobacteri
43	484	17.9	507	2 Q9EXL0	Q9exl0 streptomyc
44	483	17.8	549	2 Q93QX5	Q93qx5 corynebacte
45	482.5	17.8	544	2 Q8KP34	Q8kp34 agromyces m

ALIGNMENTS

RESULT 1

ID	Q7UF87	PRELIMINARY;	PRT;	537 AA.
AC	Q7UF87;			
DT	01-OCT-2003 (TREMELrel. 25, Created)			
DT	01-OCT-2003 (TREMELrel. 25, Last sequence update)			
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)			
DE	Phytoene dehydrogenase (EC 1.14.99.-).			
GN	CRT1 OR R810270.			
OS	Rhodopirellula baltica.			
OC	Bacteria; Planctomycetes; planctomycetacia; Planctomycetales;			
OC	Planctomycetaceae; Pirellula.			
OX	NCBI_TaxID=117;			
RN	_TaxID=117;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1;			
RX	MEDLINE=22735913; PubMed=12835416;			
RA	Gloekner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,			
RA	Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,			
RA	Schlesner H., Amann R., Reinhardt R.;			
RT	"Complete genome sequence of the marine planctomycete Pirellula sp.			
RT	strain 1."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).			
DR	EMBL; BX294151; CAD78796.1; -			
KW	Oxidoreductase; Complete proteome.			
SQ	SEQUENCE 537 AA; 60136 MW; A16C389C746949B6 CRC64;			

Query Match 37.2%; Score 1009; DB 16; Length 537;

Best Local Similarity 41.5%; Pred. No. 1.3e+68;

Matches 204; Conservative 97; Mismatches 181; Indels 10; Gaps 6;

QY	5	KH:ITVAGPGGLCAGMLLSQRFKVSIFDKHAEIGGRNPINNNGTFTGPTFLMKG 64
DB	25	KKVWVGAGPGGLASAMQLAAGGCDVTILERGGVGGRGRTSAIEDGFRFCGPTFFLYPR 84

QY	65	VLDENFELCERSSELYFLP---LSPWRYLLY-DDRDIFVYSDRNNRAELQSVFDEGT 120
DB	85	VLDIEFHSTGR---DLMEKRVPMERLDPOYRLTFGGGGGLDCTPMDENDMDRQIAQSPQDV 141

[illegible]

GO: GO:0016491; P:oxidoreductase activity; IEA.
GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR000759; Adrmdx_reductase.
DR PRINTS: PR00419; ADXRDPTASE.
KW Oxidoreductase; Complete proteome.
KW SEQUENCE 498 AA; 56453 MW; C12B1FF421DEB31B CRC64;

Query Match 29.8%; Score 808; DB 16; Length 498;
Best Local Similarity 34.4%; Pred.No.2.8e-53;
Matches 168; Conservative 109; Mismatches 504; Indels 8; Gaps 7;

2Y 5 KHIIVAGAGCGLCACMLLSQRGFKYISFDKHAIEIGRRNRPINMGFTDPTGTFLLMKG 64
2 KQVSIIGAGIGLTAARVLQKLGYNVTIYEQAQPGKQNIQVSQGFIDVGEFTVMKD 61
65 VLDWMFELCERSSDYLEPLPLSPMTHLLYDDRDIIFYVS-DRENWRAELQRVFDSTGY 123
62 IYERVFRFCGVDPENYLPFFBEVQPLMELVEGDSSLDLSRDLPTLIAENRIAPDVMGM 121
124 EQFEQBRKRKNALYP-CITRDYSSLSKFSLSLDLIXALPWLAPPKSVFNNGLCQYFNQKM 182
122 LNFELADIVHRTIAKPNFLERSFNWEDFTNLALYAGLQRTFNNAKNIKFNENL 181
183 ELAFCFSQKYLGMSPWTCPALFTMLPVLEHYGIYVKGGLNRIAAQAQVTAENGGEIH 242
182 KNSLAFQTLYIGISPYQGPSLYNIPIELFYGVFLKGMVTYVNSLVKLFKEQGGTLK 241
243 LNSPESLIIENGAAKGVKLQHGAEIARGDVIINADFAHAWTHLVKQVLKK--VTPENL 300
242 LATPVQEIILIKDKVAGIRV-NNPIHSDIYLCALFPTANTYLI-PNEQDAGQYTDQKI 299
301 KQREYSCSTFMYLGLDKIKYIDLPH-HTIVFAQYDTYNNIRNIPDNKTYLDDFSYVONASA 359
300 AKWDYSCLCVLVLGLDKKYPTEALHTIHPAQDFSKXVADLPNGKLPEDPSYVYVPSK 359
360 SDDSLAPAGKSAIYLVIPMNDSGLDQK-AKQNVREQVLDTLGAELGLSDIRAHICE 418
360 WDPSLAPENSALYLVVPVPELSKFNDSPTTLKRVQLINKLRTATFSIDIEQIIVE 419
419 KIITPOTWETDEHVYKATFSLSHKFSQMLYWRPNRFEELANCYLVGGGTHPGSGLPTI 478
420 KQFTPVDFKEQFGAYGATFGLRLTLAKSNYVRPHNKEDYADHLVFCGSSITHEGAGVPIV 479
479 YESARISAK 487
480 MQSAKLAVE 488

RESULT 6
Q8YKN8 PRELIMINARY; PRT; 499 AA.
AC Q8YKN8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Zeta-carotene desaturase.
GN ALL7255.
GN Anaebaena sp. (strain PCC 7120).
OS Plasmid pCC7120alpa.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
QX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anaebaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003600; BAB7833.1; -.

[illegible]


```

R InterPro; IPR000759; Admrx_reductase.
R InterPro; IPR008151; Phyt_n_hydro.
R PRINTS; PR00419; ADXRDTASE.
R PrcDom; PD139017; Phyt_n_dehydro; 1.
Q SEQUENCE 499 AA; 56743 MW; AC1CF06EF745B4D34 CRC64;

Query Match.          28.9%; Score 783.5; DB 2; Length 499;
Best local Similarity 35.0%; Pred. No. 2.le-51;
Matches 175; Conservative 103; Mismatches 201; Indels 21; Gaps 10;

Y 4 TKHIIIVGAGPGGLCAGMLLSORGPKVSIYFDKHAIEIGGGRNRPINNMGGFTDGPFPFLMK 63
b 2 SKKVAIVGAGPGGLATAIRLAGYQVEIFEAAERVGGEMRGFEVDSVAEDTGPILQLP 61

Y 64 GVLDMEFELCERSEYLYEFLPSMYRL-LYDDRRDIFVYSRENRRAELQAVDEGTG 122
b 62 HLYKELFEEAGLAFADYVOLKLEPYTRLKFWMDGTOLDITSLQSKFKTQLATIRSDLFLA 121

Y 123 YEOFMEOGRKFNALY-PCITRDYSILKSPFLS-DLIKAL---PMLAPKSVFNMLGVF 177
b 122 FDRWYSEHIRKYLHGKPYLAGPARSIFGYLRPDDLKFLSPFW---ENLYQHWFFP 177

Y 178 NQKMRALAFCSQKILGMSPEWCPALFTMLPYLEHEYGIYHYKGLNRITAAAMAQVIEN 237
b 178 QDERLVYDLRYPKYLGMHPTVASSVFSLIPIFEFSQGVVHFPVGGFRALQAQGLANAADL 237

Y 238 GGEHLNSETESLIIENGAKGVKIQHGAELRGDEVIINADFAHAMTHLAVKPGVLKXYTP 297
b 238 GVKIHLSPVHQ-WIDQGVQRGLELADASHQGFDTVIINADFAIYVRHLLPITSAGRTID 297

Y 298 ENLKQREYSGTFMLYLGLDKIY-DLPHTTIVFAKDYTTNIRNI----FDNKTLTD-DPS 351
b 298 NKLQMQVFCSTFMLYGINRRYEDLPHQIYL---SDNIRLERPWVDDSGALDSTDP 353

Y 352 FYVQNASASDLSAPACKSALYVIVMPNRDPSGLDWOAHQVREQVILTLGARIGLSDI 411
b 354 FYVCNPTIIDFSNAPAGHSTFLVLPVTPNTSYAVDWDIKQSYTDFILKRLHL-LGYHNI 412

Y 412 RAHTECKILITPOTWETDEHHVYKGFSLSHKFSOMLYWRPHNRPEELANCYLVGGTHP 471
b 413 EQHIVTOSCTYQSWLDDYRVHLGAVNLSHNLTLQGPFPPIRSENIAGLYWIGGAVHP 472

Y 472 GSGLPITYESARISAKLIISQ 491
b 473 GSGLLTILEASRSAAGFIHQ 492

```

```

RESULT 8
GENNO PRELIMINARY; PRT; 511 AA.
C Q8ENM0
T Q8ENM0;
T 01-MAR-2003 (TReMBLrel. 23, Created)
T 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
T 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
T E Phytoene dehydrogenase (phytoene desaturase) [EC 1.3.-].
T N OB2461.
T S Oceanobacillus iheyensis.
C Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
X NCBI_X TaxID=182710;
X [1]
X SEQUENCE FROM N.A.
X STRAIN=HTE831 / DSM 14371 / JCM 11309;
X MEDLINE=22220767; PubMed=12235376;
X Takami H., Takaki Y., Uchiyama I.;
X "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
X Ridge and its unexpected adaptive capabilities to extreme
X environments.";
X Nucleic Acids Res. 30:3927-3935 (2002).
X EMBL; AP004601; BAC14417.1;
X GO; GO:00016491; F:oxidoreductase activity; IEA.
X GO; GO:0006118; P:electron transport; IEA.
X InterPro; IPR000759; Adrndx reductase.
X InterPro; IPR002937; Amino Oxidase.

```

[illegible]

```

RESULT 9
Q3KZ28      PRELIMINARY;      PR7;      514 AA.
AC      Q8KZ28;
DT      01-OCT-2002 (T:EMBLrel. 22, Created)
DT      01-OCT-2002 (T:EMBLrel. 22, Last sequence update)
DT      01-OCT-2003 (T:EMBLrel. 25, Last annotation update)
DE      Phytoene dehydrogenase.
GN      CRT1.
OS      uncultured proteobacterium.
OC      Bacteria; Proteobacteria; environmental samples.
OX      NCBI_TaxID=153809;
RN      [1]
RP      SEQUENCE FROM N. A.
RX      MEDLINE=21822632; PubMed=11832943;
RA      Beja O., Suzuki M.T., Heideberg J.P., Nelson W.C., Preston C.M.,
RA      Hamada T., Eisen J.A., Fraser C.W., Delong E.F.;
RT      "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
RL      Nature 415:630-633(2002).
DR      EMBL: AE008921; RAN48646.1; -.
DR      GO: GO:0016491; F:oxidoreductase activity; IEA.
DR      GO: GO:0009058; P:biosynthesis; IEA.
DR      GO: GO:0006118; P:electron transport; IEA.
DR      InterPro: IPR02937; Amino oxidase.
DR      InterPro: IPR008150; Bac phytoene dh.
DR      InterPro: IPR00205; NAD BS.

```

DR InterPro; IPR008151; Phyt_n dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 514 AA; 56924 MW; D5177500CAB56773 CRC64;

Query Match 26.94; Score 729; DB 2; Length 514;
Best Local Similarity 32.24; Pred. No. 3.2e-47;
Matches 160; Conservative 101; Mismatches 210; Indels 26; Gaps 7;

Y 8 IIVGAGPGGLCAGMLLSQSGFKVSIQDKHAEIGGRNRPINMGFTDTGTFTFLMKGVLD 67
Y 14 VVIGAGLGLLAAMRLKAGYATVLDKLRVGGSSVTQDGRFLDPTVTMPKVE 73
Y 68 EMFLCERRSDYLEPLSPMYALLYDDRDIFVYS--DRENMAELQRPVDEGTGVEOF 126
Y 74 SLMAACCRDHFADVDLRPLPFPFVEIRWPDGYSFRAAGDEKMQSEVQRINPADLPKYKF 133
Y 127 MEQERKFNALYPCIT-----RDYSSLSKSLSLDLKALPWLAPPKSVFNLLQYFNQEK 181
Y 134 LKDSQRYIIIGYEGWAEPMRLWETLKVLPFAMLR-----DSIYGLAARVKDER 187
Y 182 MRLAFCSQSKYLGMSPWPCPALFTMLPYLEHEYGIYHVKGGLNRIAAQAQVIAENGGEI 241
Y 188 LRMALSFHPLFIGDPHVTSTIYALVAHLEKTYGVHYAMGVOQIADAMAAVVRAGGQI 247
Y 242 HLNSEISLTIENGAAKVQLQHGAELEGRDEVIINADFAHAMTHLVKPGVLYKKTYPENLK 301
Y 248 KQNAVDAILIKNAGAAQAVLTDQREDAPLVSNADAGTYDHLNHSRRRTWTKLA 307
Y 302 QREYSCSTFMYLIG----LDKIYDLPHTTIVPAKDYTTNIRNIPDNKTLDDFSPVQNA 357
Y 308 RKRSMGLFWYFGTRCTAGRWADVGHITANGPRYKGLLRDIFLKRSLDDSLVIHRP 367
Y 358 SASDSDSLAPAGKALYLVPMNN--DSGLDWQAHCONVREQV---LDTLGARLGLSDIR 412
Y 368 SVTDPSPVAPAGDITFYVLSVPFHLGKNKVDQEMPIYAKVAAREVKL-----MEGPE 422
Y 413 AHIECEKIITQWTEDRHYVKGATFSLSHKFSQMLYWRPHNRPPELANCYLVGGGTHPG 472
Y 423 ACISTETITFETEDRYLSPHGAGFSIEPILQASAFRPHNNSSEARGLYLVGAGTHPG 482
Y 473 SGLPTIYESARISAKLI 489
Y 483 AGLPGVSSIAEVLKLV 499

RESULT 10
Q840T3 PRELIMINARY; PRT; 513 AA.
AC Q840T3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTL.
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RC Strain=SI;
RA Ouchane S.; Steunou A.-S.; Astier C.;
RT "Photosynthesis Gene Regulation in Rubrivivax gelatinosus:
RT Transcription Factor PpsR is Involved in both Negative and Positive
RT Control";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; A234385; A093135.1; -;
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.

DR InterPro; IPR008150; Bac_phytoene_dh.
DR InterPro; IPR000205; NAD_ES.
DR InterPro; IPR008151; Phyt_n dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 513 AA; 57074 MW; 9C46447BB02F682D CRC64;

Query Match 26.88; Score 726; DB 2; Length 513;
Best Local Similarity 32.64; Pred. No. 5.4e-47;
Matches 159; Conservative 97; Mismatches 208; Indels 24; Gaps 8;

QY 8 IIVGAGPGGLCAGMLLSQSGFKVSIQDKHAEIGGRNRPINMGFTDTGTFTFLMKGVLD 67
DB 24 LVVSGFGGWAAPVLAAGYRVTVLEKLDAPGGRAYVHRREGHVDFDAGPTIVTPVLPFD 83
QY 68 EMFLCERRSDYLEPLSPMYALLYDDRDIFVYS--DRENMAELQRPVDEGTGVEOF 126
DB 84 ELWALAGKTFSDDIELKSLDPFYAIRDFDDGDHFDYSGDPARMRAEVRRIISFSDAEGFERF 143
QY 127 MEQERKFNALYPCITRDYSSL--KSFLSL-DLIKALP-----WLAFFKSVFNLLQY 176
DB 144 MREADQ-----CYELGFRTEGDKAFDTVGDLLKAAPLIIRLGM-----RSLHQMVS 192
QY 177 FNQBRMLAFCSQSKYLGMSPWPCPALFTMLPYLEHEYGIYHVKGGLNRIAAQAQVIAE 236
DB 193 LKHPKLIAMSLQSLIGGNPFVSITSMYALVNALERQWGVHWAMGGTGERLIRGLVDVFEG 252
QY 237 NGCEITHLNSLTIENGAAKVQLQHGAELEGRDEVIINADFAHAMTHLVKPGVLYKKTYP 296
DB 253 MGGTWRLKAEVKRLEVNGVATGVTLADGERLPADIVVCGNDGTLYLKNLVDARWRKHT 312
QY 297 PENLKQREYSCSTFMYLIGLDKIY-DLPHTTIVPAKDYTTNIRNIPDNKTLDDFSEVVO 355
DB 313 DARTERGHYSGMLFVWYFGTDRYEDVPHNMVLGPRYRELLDDIFRKKGLAGDFSILH 372
QY 356 NASADSDSLAPAGKALYLVPMNNDSGLDWQAHCONVREQVLDITLGARLGLSDIRAH 415
DB 373 RPTATDPSPVAPAGDITFYVLSVPFHLGKNKVDQEMPIYAKVAAREVKL-ERTVLPG 431
QY 416 ECEKIITQWTEDRHYVKGATFSLSHKFSQMLYWRPHNRPPELANCYLVGGGTHPGSL 475
DB 432 KVSFCTTFLDPQHRLLSYKGAGFGLEPLLLQSAFYRPHNRSSEYKNIPLFMVGASTHPG 491
QY 476 PTIYESAR 483
DB 492 PGVIMSAR 499

RESULT 11
Q8JP98 PRELIMINARY; PRT; 511 AA.
ID Q8JP98
AC Q8JP98;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTL.
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL144;
RA Nagashima K.V.; Shimada K.; Matsuura K.;
RT "Phylogenetic analysis of photosynthetic genes of Rhodocyclus
RT gelatinosus: Possibility of horizontal gene transfer in purple
RT bacteria";
RL Photosyn. Res. 36:185-191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL144;


```

Db 361 HAPSVTDSLAPQGCSTYYVLSFVPHLAAPIDWSVEGPRVDRIIDVLEARI-LPGKLS 419
QY 414 HIECEKIITPQWETDEHYKGFATSLSHKFSQMLYWRPHNRFEELANCYLGGCTHPCS 473
Db 420 DIATCRITFDNFTELNAHLGSAFSLPILTQSAFYSAHNAADKKIKGLYLVGAGTHPGA 479
QY 474 GLPTIYESARISAKLSQKHR 494
Db 480 GIPGVVGSATARVLEDDQR 500

RESULT 13
Q47845
ID Q47845 PRELIMINARY; PRT; 492 AA.
AC Q47845;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTI.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93371414; PubMed=895826;
RA Liu S.T.;
RT "Carotenoid-biosynthesis genes as a genetic marker for the purpose of
RL gene cloning.";
RL Biochem. Biophys. Res. Commun. 195:259-263(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94236237; PubMed=8180698;
RA To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
RA Chang Y.S., Liu S.T.;
RT "Analysis of the gene cluster encoding carotenoid biosynthesis in
RL 'Erwinia herbicola' Eho13.";
RL Microbiology 140:331-339(1994).
DR EMBL; M90698; AAA21263.1; -.
DR PIR; S52586; S52586.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phyt_n dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 492 AA; 55010 MW; 2D65E1A2A32D0635 CRC64;

Query Match 26.3%; Score 713.5; DB 2; Length 492;
Best Local Similarity 31.7%; Pred. No. 4.6e-46;
Matches 158; Conservative 107; Mismatches 208; Indels 25; Gaps 10;

QY 5 KHIIIVGAGPGCLCAGMLLSQRFKVSIFDKHABIGGRNRPINNGFTTDTGTFLMKG 64
Db 2 KPTTVIGAGFGGLALAIPLQAGIPVLLLEQDKPGRVYVYDQGTTFDAGTVIDPS 61
QY 65 VLDENFPI-CERSESDYLFPLSPMYRLLYDDRIFFVY-SDRENMRALQRFVDEGTGY 123
Db 62 AIELFTLAGQLKDYVELLPVAPFVRLCWESGKVFYNDQALEAQIOCFNPRDVEGY 121
QY 124 EQFQEQRKRNFALYPCITRDYSSLK----SFLSL-DLIALPWLIA---PPKSVFNGLQ 175
Db 122 RQFLDYSRAVFEGY-----LKLGVTFLLSFRDXLRAAQLAKLQWRVTVYSKVAS 172
QY 176 YNQEMRLAFCTOSKYLGMSPWECPALFTMLPYLHEGYIHYVKGGLNIAAQAQVIA 235
Db 173 YIEDEHLQAFSFLVGGNPFATSSYIILHALEREWGVPFRGGTALVKGMIKLQ 232

```

```

QY 236 ENGGBIHLNSELISLIENGAAGKVKLQHGAEIRGDEVINADFAHAMTHLVK--PGVLK 293
Db 233 DLGGEVLNAKVSHEMTTGDITIEAVHLEDGERFPFTRAVASNADVVHTYRDLISQHPAAVK 292
QY 294 KYTPENLKQREYSCSTFMLYGLQKIYD-LPHHTIVFAKYDTTMRNIFDKNTITDDPSF 352
Db 293 Q--SKKLQTKRMSNSLFVLYFGLAHHHDLQLAHTVCFGPRYRELTHEIFNEDGLADDFSL 350
QY 353 YVONASASDDSLAPAGKSAALYVLPMPN-NDSGLDWQAHQCNVREOVLDTGARLGLSDI 411
Db 351 YLHAPCVTDSSLAPEGCGSYVLPVPLGTLTANLDWTVEGPRILDRIFEYLEQHY-MPGL 409
QY 412 RAHIECEKIITPQWETDEHYKGFATSLSHKFSQMLYWRPHNRFEELANCYLGGCTHCP 471
Db 410 RSLQVTQRMFPFQDQLNAYQGSFVSPVLTQSAWFRPHNRDKTINNLYLVGAGTHP 469
QY 472 GSGLPITYESARISAKLI 489
Db 470 GAGIFGVIGSAKATAGLM 487

RESULT 14
Q93CI7
ID Q93CI7 PRELIMINARY; PRT; 504 AA.
AC Q93CI7;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTI.
OS Xanthobacter sp. (strain Py2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Xanthobacter.
OX NCBI_TaxID=78245;
RN [1]
RP SEQUENCE FROM N.A.
RA Larsen R.A., Metcalf W.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF408849; AAL02000.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phyt_n dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_n dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 504 AA; 55556 MW; CA848D11A607A828 CRC64;

Query Match 26.0%; Score 705.5; DB 2; Length 504;
Best Local Similarity 31.4%; Pred. No. 1.9e-45;
Matches 154; Conservative 108; Mismatches 211; Indels 17; Gaps 8;

QY 9 IVGAGPGCLCAGMLLSQRFKVSIFDKHABIGGRNRPINNGFTTDTGTFLMKGVLD 68
Db 14 VIGSGFGSLAIRLQAAGIRTTVEQDKPGRVYVYDQGTTFDQGTPTVIDPTCLB 73
QY 69 MEELCERSESDYLFPLSPMYRLLYDDRIFFVY-SDRENMRALQRFVDEGTGYEQPM 127
Db 74 VFAAGRKLSDYVELLPVAPFVRLCWESGKVFYNDQALEAQIATFDKADVEGYRFL 133
QY 128 EQERKRNFALY-----PCITRDYSSLKSLFLSLDLKALPWLAPPKSVFNGLQYFNQEX 181
Db 134 AYSQAVFEGYKLGAVPFL--EPSSMMK-AAPKLVRLEAW---RSVYAVSRFRIDDD 186
QY 182 MRLAFCOSKYLGMSPWECPALFTMLPYLHEGYIHYVKGGLNIAAQAQVIAENGGEI 241
Db 187 LRQAFPSFLVGGNPFATSSYIILHALEREWGVPFRGGTALVKGMIKLQFDLQGEI 246
QY 242 HLNSBIESLIENGAAGKVKLQHGAEIRGDEVINADFAHAMTHLVKPGVLKRYTPENLK 301

```

b 247 RLTPVDEIVVEQRATAVKLKSGETLPLFDLVASNADVVHTYRHLRGARSGEGARIA 306
Y 302 QREYSCSTFMYLGLDKIYD-LPHHTIIVPAKDYTTNIRNIFUNKTLTDDPSFYVONASAS 360
b 307 KTRHNSLSFTYFGARITWDHLQHTVLFPGPRYRGLVDEIFKGPMLPDDFSLYLHAPT 366
Y 361 DDLAPAGKASALYVLVPMFN-NDSGLDQAHQCNVREOVLDFTIGARLGLSDIRAHIECEK 419
b 367 DKSLAPEGCTAFVLSFVPHLKGADIDMEVEGPLYRDIRLAHLEBRL-LPGLRSLVTSR 425
Y 420 ITPQWETDEHYVKGATSLSHKFSQMLYWRPHRFEELANCYLVGSGTHPGSLPTIY 479
b 426 ILTPFGFRDELSAHQSAFSVEPLLTQSAWFRPHNRDAKIANLYTAGAGTHPGAGVPGW 485
Y 480 ESARISAKLI 489
b 486 GSAKATAGLI 495

RESULT 15

8GCS1 PRELIMINARY; PRT; 492 AA.
C 08GCS1;
T 01-MAR-2003 (T-EMBLrel. 23, Created)
T 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Phytoene desaturase.
N CRTI.
S Pantoea stewartii.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=66289;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 8200;
A deSouza M.L., Kollmann S.R., Schroeder W.A.;
L "Carotenoid Biosynthesis (MO 02/079395 A2).";
T Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AY166713; AA85559.1; .
R GO: 0016491; F:oxidoreductase activity; IEA.
R GO: 0009058; P:biosynthesis; IEA.
R GO: 0006118; P:electron transport; IEA.
R InterPro: IPR002937; Amino oxidase.
R InterPro: IPR008150; Bac phytoene_dh.
R Pfam: PF01593; Amino oxidase, 1.
R PROSITE: PS00982; PHYTOENE_DH; 1.
Q SEQUENCE 492 AA; 54836 MW; F3DCD224547A5FEA CRC64;

Query Match 25.9%; Score 703.5; DB 2; Length 492;
Best Local Similarity 31.7%; Pred. No. 2.7e-45;
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;
Y 5 KHIIIVGAGGGGLCAGMLLSQRFKYSIDFKHAEICGRNRPINMGFTDTGPTFLMKG 64
b 2 KPTTVIGAGGGGLALAIRQAQIPVLLLEQDKPGRAYVVOEGFTFDAGPTVITDPS 61
Y 65 VLDEMPFLCRRRDELYLEPLSPMYRLYYDDRDIFVY-SDRENHAEIQRVDEGTDGY 123
b 62 AIELFALACKQKQVYELLPTVFPYRLCWESKVFYNDQQLQEAQIQFNPRDVAGY 121
Y 124 EQFMEQERKFNALYPCITRDYSLK-----SFLSL-DLIKALFWLA---FPKSVFNILQQ 175
b 122 RAFLDYGRAVFNQY-----LKLGTVPFLSFKDMLRAAPQLAKQAWRSVYSKVAG 172
Y 176 YFNQEKRLAFQCSKYLGMSWPECALPTMLPYLEHYGIYHVGGLNRIAAQAQVIA 235
b 173 YIEDEHLRQAFPSHLLVGGNPFATSIYFLIALERENGVWPPRGCTGALVNGAKIKLFQ 232
Y 236 ENGGEIHLNSEIIEIENGAKGVKIQHGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
b 233 DLGGEVVLNARVSHMETVGDKIQAQVLEDGRRETCVASNADVVHTYRDLLSQHPAAK 292
Y 294 KYTPENLKQREYSCSTFMYLGLDKIYD-LPHHTIIVPAKDYTTNIRNIFUNKTLTDDPSF 352

Db 293 Q--AKKLGKRMNSLFLYFGLNHHHQLAHTVCFGPRYRELTHEIFNHDGLAEDFSL 350
QY 353 YVQNASASDDSLAPACKSALYVLVPMFN-NDSGLDQAHQCNVREOVLDFTIGARLGLSDI 411
Db 351 YLHAPCVTDPSLAPEGCGSYVLAAPVPHLGTANLDNAVEGPRLRDRIFDYLEQHY-MPGL 409
QY 412 RAHIECEKLIITPQTWETDEHYVKGATSLSHKFSQMLYWRPHRFEELANCYLVGSGTHP 471
Db 410 RSQLVTHRMFTPFDFRDELNAMQGSFAFSVEPLLTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
QY 472 GSGLPTIYBSARISAKLI 489
Db 470 GAGIPGVIGSAKATAGLI 487

Search completed: February 29, 2004, 14:50:43
Job time : 49.0179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 8.61236 Seconds
(without alignments)
3089.496 Million cell updates/sec

itle: US-09-941-947A-22

fect score: 2711

quence: 1 WANTKHIIIVAGPGGLGAG.....KRVRFKDIHSAWLKXKA 511

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

tal number of hits satisfying chosen parameters: 141681

nimum DB seq length: 0

ximum DB seq length: 200000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	825.5	30.5	529	1	CRTI_MYXXA
2	794	29.3	517	1	CRTU_MYXXA
3	743	27.4	524	1	CRTI_RHOCA
4	737	27.2	518	1	CRTI_RHOSH
5	704.5	26.0	492	1	CRTI_ERWHE
6	700.5	25.8	492	1	CRTI_PANAN
7	686	25.3	583	1	CRTI_PHYEL
8	681	25.1	595	1	CRTI_NEUCR
9	607	22.4	621	1	CRTI_CERNC
10	605.5	22.3	501	1	CRTI_AGRAU
11	508	18.7	507	1	CRTI_STRGR
12	493	18.2	508	1	CRTI_STRGR
13	326	12.0	495	1	CRTD_RHOSH
14	325	12.0	494	1	CRTD_RHOCA
15	175	6.5	469	1	P49_STRLI
16	139	5.1	472	1	CRTI_SYNY3
17	138.5	5.1	393	1	YF88_METVA
18	134	4.9	500	1	PC11_ARATH
19	129.5	4.8	558	1	ZDS_ARATH
20	128.5	4.7	570	1	CRTI_SOYVN
21	126.5	4.7	516	1	OXLA_AGRKH
22	126.5	4.7	588	1	ZDS_CAPAN
23	125.5	4.6	587	1	ZDS_TARER
24	123	4.5	516	1	OXLA_CROAD
25	120.5	4.4	471	1	GLTD_ECOLI
26	119.5	4.4	588	1	ZDS_LYCES
27	118.5	4.4	583	1	CRTI_LYCES
28	117.5	4.3	445	1	GDIC_RAT
29	117	4.3	749	1	TR2M_AGRRH
30	115.5	4.3	447	1	GDIA_HUMAN
31	115.5	4.3	447	1	GDIA_PANTR
32	115.5	4.3	582	1	CRTI_CAPAN
33	115	4.2	566	1	CRTI_ARATH

34	114.5	4.2	445	1	GDIC_MOUSE
35	114.5	4.2	447	1	GDIA_CANFA
36	114.5	4.2	447	1	GDIA_MOUSE
37	114.5	4.2	447	1	GDIA_RAT
38	114.5	4.2	613	1	YC68_CAMJE
39	112.5	4.1	445	1	GDIB_HUMAN
40	112.5	4.1	447	1	GDIA_BOVIN
41	112.5	4.1	639	1	YGFT_ECOLI
42	111	4.1	471	1	PF0X_MYXXA
43	111	4.1	508	1	FMS1_YEAST
44	109.5	4.0	566	1	CRTI_ORYSA
45	109.5	4.0	574	1	ZDS_ARPS

ALIGNMENTS

RESULT 1

CRTI_MYXXA
ID CRTI_MYXXA STANDARD; PRT; 529 AA.

AC Q02861;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CARC.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxId=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR403;
RX MEDLINE=93223667; PubMed=8467787;
RA Fontes M., Ruiz-Vazquez R.M., Murillo F.J.;
RT "Growth phase dependence of the activation of a bacterial gene for
RT carotenoid synthesis by blue light."
RL EMBO J. 12:1265-1275(1993).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
CC intermediaries of phytofluene, zeta-carotene and neurosporene by
CC the introduction of four double bonds (By similarity).
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- INDUCTION: By blue light.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94727; AAA25390.1; -
CC PIR; S35306; S35306.
CC InterPro; IPR000759; Adrmdx reductase.
CC InterPro; IPR002937; Amino oxidase.
CC InterPro; IPR008150; Bac phytoene dh.
CC InterPro; IPR000205; NAD BS.
CC InterPro; IPR008151; Phytn dehydro.
CC Pfam; PF01593; Amino oxidase; 1.
CC PRINTS; PR00419; ADXDRASE.
CC ProDom; PD139017; Phytn dehydro; 1.
CC PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 529 AA; 53536A8DFD0D24BC CRC64;

Query Match 30.5%; Score 825.5; DB 1; Length 529;

Best Local Similarity 35.1%; Pred. No. 2.1e-54;

Matches 179; Conservative 106; Mismatches 192; Indels 33; Gaps 10;

3 NTXHIIVGAGPGGOLCAGMLLSQGRGKVSIFDKHAEIGGRNRPINMNG---FTFDGTGTF 59
7 SVRHVIVGAGPGGLSAALNLAGQFRVTVVXDVAVPGGRMKGLTLGASGEYAVDTGPSI 66
60 LLMKGLDENFELCERSEBYLFLPLSPMYRL-LVDDRDI FVYSRENKRAELQORVEDE 118
67 LQPGVLEQIFRRAARLEDYVKLLPDVNVTRVHFWDGTHLDTTRHLDKMEALAKFGR 126
119 GTDGYEFMEQERKFNALYP---CITRDYSSLSKFLSLDLIKALPWLAPP-----KS 168
127 QASALRQWMDGREGKYGIAVQKEICTSAD-----NLGYVAPWRLAPTLLRKFPMQT 176
169 VFNILGOYFNOEKWRLAFCQSKYLQMSWPCPALFTMLPYLEHEYGIYHVGKGLNRINA 228
177 LYRQLDGFHDDRVYALAYPSKYLGHPTTCSVFSVPFLFALFVHVGGRFELSR 236
229 AMAQVIAENGSEIHLNIEBLSIENGAAGVQLQCAELRGDEVIINADFAHAMTHLV- 287
237 GMRCARDLGNFMGTPEVKRVUDAGAVGVKLGVGEVLADAVVVDADVAARSLLP 296
288 ---KPGVLYKXYTPENLKORVYSTFMVLYGLDKIY-DLPHHTIVFAKYDTNIRNIFDN 343
297 AEAREG--SRUTDAALERAKVYSCSTFMAYVGLDTVYADLPHLLIYLSGARRTDRDAED 354
344 KTL--TDDFSFVQNASADDSLAPAGKSAIYLVPMNPNDGSLDQAHQCNVREOVLDTL 402
355 RHVLEDPPTVFCVPGVDFSGAPAGHSTLYLVPTNGRPPVWVKTEQALRERIPAML 414
403 GARLGLSDIRAHIECEKIITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELANC 462
415 -EKVGLKGVREHIREERYFYAETWRDENVFRGAVFNLSHTWQLGLPLPKVKNRDIEGL 473
463 YLVGGTGRSGGLPTIYESARISKLISQK 492
474 YFVGCGTHPGSGGLTINESANIAADYLTRE 503

RESULT 2

RTJ MYXA
ID CRTJ MYXA STANDARD; PRT; 517 AA.
AC PS4979;
JT 01-OCT-1996 (Rel. 34, Created)
JT 01-OCT-1996 (Rel. 34, Last sequence update)
JT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
ZN CARA2
XS Myxococcus xanthus.
XC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
XC Cystobacteriineae; Myxococcaceae; Myxococcus.
XC NCBI_TaxID=34;
ZN [1]
ZF SEQUENCE FROM N.A.
XC STRAIN=DK1050;
XA MEDLINE=96061955, PubMed=7588751;
XA Botella J.A., Murillo F.J., Ruiz-Varquez R.M.;
XA "A cluster of structural and regulatory genes for light-induced
XA carotenogenesis in Myxococcus xanthus.";
XL Eur. J. Biochem. 233:238-248(1995).
XC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
XC intermediary of phytofluene by the symmetrical introduction of two
XC double bonds at the C-11 and C-11' positions of phytoene.
XC -!- COFACTOR: FAD (Probable).
XC -!- PATHWAY: Carotenoid biosynthesis.
XC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
XC -----
XC This SWISS-PROT entry is copyright. It is produced through a collaboration
XC between the Swiss Institute of Bioinformatics and the EMBL outstation -
XC the European Bioinformatics Institute. There are no restrictions on its
XC use by non-profit institutions as long as its content is in no way
XC modified and this statement is not removed. Usage by and for commercial
XC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
XC or send an email to license@isb-sib.ch).
XC -----

DR EMBL; Z21955; CAA79956.1; -.
DR PIR; S32169; S32169.
DR InterPro; IPR000759; Admrx reductase.
DR InterPro; IPR002937; Amino Oxidase.
DR InterPro; IPR008150; Bac.phytoene_dh.
DR InterPro; IPR008205; NAD_BS
DR InterPro; IPR008151; Phytin_dehydro.
DR Pfam; PF01593; Amino_oxidase; 1.
DR PRINTS; PR00419; ADXRDITASE.
DR PRODOM; PD139017; Phytin dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 11 44 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 517 AA; 2793 MW; CECDD74AB9F9F8CB CRC64;
Query Match 29.3%; Score 794; DB 1; Length 517;
Best local similarity 35.8%; Pred. No. 4.7e-52;
Matches 177; Conservative 106; Mismatches 201; Indels 10; Gaps 8;
QY 5 KHIIVGAGPGGOLCAGMLLSQGRGKVSIFDKHAEIGGRNRPINMNGFTFDGTGTFLLMKG 64
DB 8 RRVVVGAGVGLAAARLAHQGFVQFSEKTPGRCNRLQVDTWLDGFTVILMPE 67
QY 65 VLDMFELCERRSEDYLFELSPMYRLLYDDR-DIEVYSDRENKRAELQORVEDEGTGY 123
DB 68 VPETFRVAGGRIEDYLTLLRCDPNRYRVHFRDRSDVTFTSELCAVGRLELERVEFGSVARY 127
QY 124 EQFMEQERKFP-NALYPCITEDYSSLSKFLSLDLIKALPWLAPKSVFNILGOYFNQBM 182
DB 128 LAFLAQGRVQTRTSLDLHVGRTAGLDYLSPLVARIPOVRAHRRMYADVSFFQDERL 187
QY 183 RLAPCFOSKYLQMSWPCPALFTMLPYLEHEYGIYHVGKGLNRINAAMQVIAENGEITH 242
DB 188 RAAMFTQVLYGVSPYASPAVYGLLPTELGVGIWPKGLGYAIPOALERLAREREGVRFH 247
QY 243 LNSEIRSLIITENGAAKGVKLQHGAEIRGDEVIINADFAHAMTHLVKGVLYKVTPELNKQ 302
DB 248 YGAPVERILDTGRTGRVLEGGEVWEADAVLCNALPYAYEKLLDP---KATTLKRRKEK 304
QY 303 REYSCSTFMVLYGLDKIY-DLPHHTIVFAKYDTNIRNIFDNKTLTDDFSFYVQNASAD 361
DB 305 LRYTSSGVMVLYGMKRYPELLHNHNVFGEDYKGSFDDIFEFR-VPEDSPFYVNAPTRD 363
QY 362 DSLAPAGKSAIYLVPMNPNDGSLDQAHQCNVREOVLDTLGLRLGLSDIRAHIECE-KI 420
DB 364 ASLAPGKDALYLVVPVPHQHPDLWKVGPVKRAKFFARM-AELGFPFSLSDIEVERES 422
QY 421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELANCYLVGCGTHPGSGGLPTIYE 480
DB 423 STPDDWAGTNLARGSGFGLSQNFTQIGPFRPSNQDARVKNLFFVGASTQPGTGLPTVLI 482
QY 481 SAR-ISAKLISQKH 493
DB 483 SARLVTERLMTWAH 496
RESULT 3
CRTJ RHOC
ID CRTJ RHOC STANDARD; PRT; 524 AA.
AC P17054;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTJ.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003 / St Louis, and BEC404;
RX MEDLINE=89313663; PubMed=2747617;


```

[3]
RN   SEQUENCE FROM N.A.
RP   RP
RC   STRAIN-ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX   MEDLINE=20115911; PubMed=10648776;
RA   Choudhary M., Kaplan S.;
RT   "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT   sphaeroides 2.4.1.";
RL   Nucleic Acids Res. 28:862-867(2000).
CC   !- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC   intermediary of phytofluene by the symmetrical introduction of two
CC   double bonds at the C-11 and C-11' positions of phytoene.
CC   !- COFACTOR: FAD [Probable].
CC   !- PATHWAY: Carotenoid biosynthesis.
CC   !- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC   This SWISS-PROT entry is copyright. It is produced through a collabora

```



```

between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licens@isb-sib.ch).
-----
EMBL; S71770; AB311138.1; -.
EMSL; AJ010302; CAB38739.1; -.
EMSL; AF195122; AAF24289.1; -.
PIR; S49620; S49620.
PIR; T50745; T50745.
InterPro; IPRO02937; Amido oxidase.
InterPro; IPRO08150; Bac_phytoene_dh.
InterPro; IPRO00205; NAD_BS.
InterPro; IPRO08151; Phytin_dehydro.
Pfam; PF01593; Amido oxidase; 1.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 14 47 FAD (ADP PART) (POTENTIAL).
CONFLICT 73 73 R -> C (IN REF. 3).
CONFLICT 174 174 T -> S (IN REF. 3).
CONFLICT 292 292 L -> P (IN REF. 3).
CONFLICT 395 395 Q -> F (IN REF. 3).
SEQUENCE 518 AA; 57244 MW; 42383515639BP8P1 CRC64;
Query March 27.2%; Score 737; DB 1; Length 518;
Best Local Similarity 33.5%; Pred.No.8.9e-48;
Matches 170; Conservative 97; Mismatches 207; Indels 34; Gaps 10;
Y      8 IIVGAGPGGLCAGMLLSQRGFYSIPDKHAEITGGRRPINMNGFTTGTGPTFLLMKGVLID 67
       :|::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
D      14 LVISGLGGLAAAMELGNKQWVTVTDKLDVPGRGSSITQEGHREFDLGPTIVTPQSIR 73
Y      68 EMFLCERRSEDYLEPLSPMTRLLYDDRDIF-VYSDRNMRALQRPVDGDTGTEQF 126
       ::||:|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      74 DLWTCGRDFDADVELKPDPFYEVFWPPGSHFTTYRQSTEAKFAVARLSPGDVAGYEKF 133
Y      127 MEQERKEFNALPCIITRDYSLSKLFSLL-DLKALP---WLAFPKSVPNNLGQVFNOEKM 182
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      134 LKDSKKRYWTGYEDLGR-----RSMKMLNDLIKVLPTFGMRADRTVYQHAAALVKDERL 188
Y      183 RLAFCFOSKYLGMSWPCEPALFTMLPYLEBEHYGIYHVKGGLNRIAAAMAQVIAENGGEIH 242
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      189 RWALSFPHLFIGGDPPNVNTSYLVLSOLEKEFGVHYIGGVAATAAAMAKVIEQGGSFR 248
Y      243 LNSIESLIITENGAQVKLOHGAEIARGDEVIINADFHAMTHLVKPGVLKYYTPENLKQ 302
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      249 MNTVEDELVEKGATGVRLASGEVLRAGLVSNADAGTYMRLRNPRRWTDADAVIS 308
Y      303 REYSCSTFMILYLGLDKTY----DLPHHTIVFANDYTTIMIRINFDNKTLTDDFSFYQNAS 358
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      309 RMWSMGLFVMYTGTKGMWPDVGHTTIVNAPRYGLVEDIFLKGKLAKMSLYIHRPS 368
Y      359 ASDDSLAPAGKSAIYLVPMFN--NDSGLDQWAHQNTREQVLDTL-----GARGLIS 409
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      369 ITDPTVAPGEDDTFYALSPVPHLKQAQPVDQWAAPETRESYLEVLEQSMFGIGERTGPS 428
Y      410 DIRAHICEKIITPQTWETDEHYVK-GATFSLSHKPSQLYWRPNRFEELANCYLVGGG 468
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      429 -----LVFTPETFR-DRIYLSPWGAGFSTEPILQSAWFRPNISEEVANLFIVGAG 478
Y      469 THPGSGLPTVESARISAKLSQKHVR 496
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
D      479 THPGAGVPGVIGSAYEVMAKLAPAPRAR 506
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
RESULT 5
RTTI_ERWHE STANDARD; PRT; 492 AA.
ID CRTI_ERWHE
AC P22871;

```

```

DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase [EC 1.14.99.-] (Phytoene desaturase).
DE CRT1.
OS Erwinia herbicola.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OS Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549,
GN [1].
RN SEQUENCE FROM N.A.
RC STRAIN=EHO10;
RX MEDLINE=91088634; PubMed=2263648;
RA Armstrong G.A., Alberti M., Hearst J.R.;
RT "Conserved enzymes mediate the early reactions of carotenoid
RT biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
CC intermediaries of phytofluene, zeta-carotene and neurosporene by
CC the introduction of four double bonds.
CC -!- COFACTOR: PAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38423; AAA24820.1; -;
DR EMBL; M87280; AAA64981.1; -;
DR PIR; A39273; A33120.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac.phytoene dh.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF01593; Amino_oxidase; 1.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
DR Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 5 38 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8EDC5DB1562083F2 CRC64;
Query Match 26.0%; Score 704.5; DB 1; Length 492;
Best Local Similarity 31.7%; Pred. No. 2.3e-45;
Matches 158; Conservative 102; Mismatches 213; Indels 25; Gaps 9;
QY 5 KHTIIVGAPGGILCAGKLLSQSGKVSIFDKHAEIGGRNRPINMGFTDTGTFPLLMKG 64
DB 2 KKTIVGAGFGGLALAIRLQAAGIPTVLLERQDKPGRAYVWHQDQFTFDAGPTVITDPT 61
QY 65 VLDEMFELCRRSEDYLEPLPSPMYRLLYYDRDIFVY-SDRENMAELQRVEDECTDGY 123
DB 62 ALBALPTLAGRMEDYVRLLVPKFPYRLCWESKTLIDYANDSAELAEQITQFNPRDVEGY 121
QY 124 EQFWEQERKRNALYPCITRDYSSLLKS--FLSL-DLIKALP-----WLAFPKSVFNLL 173
DB 122 RRLFVLSQAVQ-----EGYLRIGSVFPLSPEDMLRAGPQLKLQAW----QSVVQSV 170
QY 174 GQYVNOEKMRAPCFOSKYLGMSPWCPCALFTMLPYIEHYGIYFVKGGLNRFAAMAQV 233
DB 171 SRFIEDHLRQAFSFTSLVGGNPFTTSIIYTLIHALEREWGVPFEGSGALVNGMKVL 230
QY 234 IANGGEIHLNSIESLIENGAAKGVKLQHGAEALRGDEVIIINADFAHAMTHLVKPGVLK 293
DB 231 FTDLGGEIELNARVEELVVDNRVSVQLRADGRI FDTDAVASNADVVNTYKLLGHHPVG 290
QY 294 KYTFENLKQREYSCSTFMLYLGLDKIY-DLPHTTIYFAKDYTTNINIFONKTLITDDFSP 352
DB 291 QKGAALAEKSMNSLFLVYFGLNQHSLQAHRTICFGRYRELIDEIFTGSALADPFL 350

```

RESULT 5
RTI_ERWT
ID_CRT
AC P22

01-FEB-1996 (Rel. 33, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
PDH1.
Cercospora nicotianae.
Eukaryota; Fungi; Ascomycota; Pezizomycotina;
Dothideomycetes et Chaetothyrionycetes incertae sedis;
Mycosphaerellaceae; Chaetothyrionycetes; Cercospora.
NCBI_TaxID=29003;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 18366;
MEDLINE=94368091; PubMed=8085820;
Zehrfahrt M., Daub M.E.;
"Isolation, sequence, and characterization of the Cercospora
nicotianae phytoene dehydrogenase gene";
Appl. Environ. Microbiol. 60:2766-2771(1994).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; U03903; AAB86388.1; --
PIR; T48646; T48646.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR008151; Phyt_dhydro.
ProDom; PD139017; Phyt_dhydro; 2.
PROSITE; PS00382; PHYTOENE DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
Transmembrane.
NP_BIND 11 44 FAD (ADP PART) (POTENTIAL).
TRANSMEM 536 556 POTENTIAL.
SEQUENCE 621 AA; 59529 MW; 6189EA7784963C8A CRC64;
Query Match 22.4%; Score 607; DB 1; Length 621;
Best Local Similarity 27.2%; Pred. No. 6.5e-38;
Matches 151; Conservative 114; Mismatches 211; Indels 80; Gaps 11;
8 IIVGAGPGGLCAGMLLSQSGKYSIFDKHAEIGGRNRPINWNGFTTGTFTFLAKGVLD 67
11 IIVGSGVGVSTAAKLARAGFHVTVLEKQNTGRCSLIHHEGYRFDQGSLLALPLGFH 70
68 EMF-ELCERRSEDYLEFLPLSPMYRLAYDDRDI-F-VYSDRENMAELQRFDEGTGGEQ 125
71 RTFAELGTSLEQGVKLLKCPNMYIHFSDEKFTLSDSLVMKTEVKW--EGKGYTR 128
126 FMEQERKFN-----ALYCIPTRDYSSLSKLSFLSLDLIKALPLWAPPKSVFNGLQYFNQEK 181
129 YLEFLKESGHYELSVREVLNREGEFTAMLRPFRLHLLQLHPFESINTRAGKYFWTER 188
182 MRLAFCFOSKYLGNMPCPALFTWMLPYLEHYGYHVKGGLNRIAAQAQVLAENGGEI 241
189 LERVFTFOSMYNGMSPDPAGTYSILQYTELAEGLIWPVGVGFRHWEALVKIGERGVDF 248
242 HUNSEIESLII--ENGAAGKVLQHGAEIRGDEVIINADPAHAMTHLVKPGVLKCKYTPEN 299
249 RMEATVKKILLSDEGVAKGVELEDGRLEADVVMNSDLVAYAEKLLP---IKTPYAES 305
300 LKQREYSCSTFWLYGLDK-LVD-PHTITVFAKDYTTNIRNIDNKTLTDPSFYQNAS 358
306 LKRGCGSCSSISFYWALDRQVPELAHNIFLADEYRESFDSIFKKHLIPDEPSFYVNP 365
359 ASDSLAPAGKSAIYVLVPM-----PNNSGL----- 385

Db 366 RVDSTAAPRGKDSVVVLPVGVHLLLEDRHASQAHQLSASRNHGHISSASPPDQGLTPTEK 425
Qy 386 -DQAHQCNVREQVLDLTGARGLGLSDIRAHIECEKLIITPQWETDEHYVKGATFSLSHKF 444
Db 426 QDPMWISLARKTILSTIQSRNV-DLTPLIHSTNSPLSMKQTFNDRGAILLSHSP 484
Qy 445 SQMLYWRP-----HNRFEELANCYLVGGGTHFGS 473
Db 485 FNVLCFRPTTRAKPGAFDAQLLKFGVLGRAAEVIIDAFGRGKDKGLGYVYGASAPGT 544
Qy 474 GLPTIYESARISAKLI 489
Db 545 GVPICLAGGALVARGI 560
RESULT 10
CRTI AGRAU
ID -CRTI AGRAU STANDARD; PRT; 501 AA.
AC P54978;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=44155;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=96062243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,
RA Ohtani T., Miki W.;
RT "Structure and functional analysis of a marine bacterial carotenoid
biosynthesis gene cluster and astaxanthin biosynthetic pathway
proposed at the gene level";
RL J. Bacteriol. 177:6575-6584(1995).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
intermediaries of phytofluene, zeta-carotene and neurosporene by
the introduction of four double bonds (By similarity).
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
biosynthetic pathway.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; D58420; BAA09594.1; --
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac_phytoene dh.
DR InterPro; IPR008151; Phyt_dhydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phyt_dhydro; 1.
DR PROSITE; PS00382; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 501 AA; 54806 MW; 5F251AF11D679358 CRC64;
Query Match 22.3%; Score 605.5; DB 1; Length 501;
Best Local Similarity 28.6%; Pred. No. 6.3e-38;
Matches 141; Conservative 110; Mismatches 227; Indels 15; Gaps 8;
Qy 5 KHIIVGAGPGGLCAGMLLSQSGKYSIFDKHAEIGGRNRPINWNGFTTGTFTFLAKMG 64
Db 9 KTAIVIGAGFGGLALAIRLQSGIATTLVLEARDFKGRAYVYHDDQGLHFDAGPTVITDPD 68

```

65 VLDEMFCERSESDYLFPLSPMYRLLYDDRDIFVY-SDRENMRARLQRFDECTDGY 123
66 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 ALKELWALTQDMARDVTLMBVSP-YRLMPPGKVFDYVNEADQLERQIAQFPNDOLGY 128
70 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 EORFEQERKRNALYPCITRDYSSLSKFLSL-DLIKALPWL-APPKSVFNILQOYFN 178
125 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 RRFIDYAEVYQBGYVKL-----GTVPFLKLGQMLKAAPALMKLEAY-KSVHAKVATFIK 182
130 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
179 QEKRLAFQFSKYLGMSPWCEPALFTMLPYLEHEYGIYHVKGGLNRIRIAAQAQVIAENG 238
180 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 DPLYLRQAFSYHTLLVGGNPFSTSIYALNEALERRGGWFAKGGTNQLVAGMVALFERLG 242
184 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
239 GEIHNSIESLIIENGNAKVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYKYTFE 298
240 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 GQMLLNARAKVIDTGPRTAGTVTLADGRALTADMASVGDVWNYRDLTGHTARGOSRAK 302
244 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 NLKQREYSCFTFMYLGL-DKIYDLPHTTVFAKDYTNIRNIFDNKTLTDFFSFVQNA 357
300 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
303 SLNAKRWMSLVFLHFGLEAPKDVAAHTILFGPRYKELVNEIFKPKLAEDFSLVHSP 362
304 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
358 SASDDSLAPACKSALYLVMPFN-NDGSLDWAQHCQNVREQVLDTLGARLGLSDIRAHIE 416
359 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363 CTTDPEWAPQGMSTHYVLAPVPHLGRADIDWAVEGPRYADRIILASLEERL-IPNLRSANIT 421
364 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 CEKIITPOTWETDSHVYKGTATFSLSHKFSQMLYWRPHNRPEELANCYLVGGGTHPGSLP 476
418 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 TTRIFTSDFASLNAHGSAPFVEPILTQSAMFRPHNRKXTIRNYLVGAGTHPGGIP 481
423 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
477 TIYESARISAKLI 489
478 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
482 GVWGSAAQAQVM 494

```

```

RESULT 11
RTI_STRGR STANDARD; PRT; 507 AA.
AC P54981; P72447;
AC 01-OCT-1996 (Rel. 34, Created)
AC 01-OCT-1996 (Rel. 34, Last sequence update)
AC 28-FEB-2003 (Rel. 41, Last annotation update)
AC Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
AC CRTI OR CRTI.
AC Streptomyces griseus.
AC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
AC Streptomycineae; Streptomycetaceae; Streptomyces.
AC NCBI_TaxID=1311;
AC [1]
AC SEQUENCE FROM N.A.
AC STRAIN=JA3333;
AC MEDLINE=97074881; PubMed=9917308;
AC Schumann G., Nurnberger H., Sandmann G., Kruegel H.J.;
AC "Activation and analysis of cryptic crt genes for carotenoid
AC biosynthesis from Streptomyces griseus.";
AC Mol. Gen. Genet. 252:658-666(1996).
AC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
AC intermediary of phytofluene by the symmetrical introduction of two
AC double bonds at the C-11 and C-11' positions of phytoene.
AC -!- COFACTOR: FAD (probable).
AC -!- PATHWAY: Carotenoid biosynthesis.
AC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
AC This SWISS-PROT entry is copyright. It is produced through a collaboration
AC between the Swiss Institute of Bioinformatics and the EMBL Outstation
AC the European Bioinformatics Institute. There are no restrictions on its
AC use by non-profit institutions as long as its content is in no way
AC modified and this statement is not removed. Usage by and for commercial
AC entities requires a license agreement (see http://www.isb-sib.ch/announce/
AC or send an email to license@isb-sib.ch).
AC -----
AC EMBL; L37405; AAA1950.1; -
AC EMBL; X95596; CAA64850.1; -
AC InterPro; IPR000759; Adnrxn_reductase.
AC InterPro; IPR002937; Amino_oxidase.

```

```

DR InterPro; IPR008150; Bac_Phytoene_dh.
DR InterPro; IPR000205; NAD_Bs.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF01593; Amino_oxidase; 1.
DR PRINTS; PR00419; ADXEDTASE.
DR PRODOM; PD139017; Phyt_n_dehydro; 1.
DR PROSITE; PS00982; OXIDOREDUCTASE; FAD; Flavoprotein; NAD.
DR Carotenoid biosynthesis; OXIDOREDUCTASE; FAD (ADP PART) (POTENTIAL).
DR NP_BIND 12 45
DR SEQUENCE 507 AA; 54509 MW; FBB97E7FE696B2AC CRC64;

Query Match 18.78; Score 508; DB 1; Length 507;
Best Local Similarity 28.08; Pred. No. 1.3e-30; Indels 64; Gaps 14;
Matches 144; Conservative 82; Mismatches 225;

QY 6 HIIVGAPGGLCAGMLLSQRFKVSIFDKHAEIGGRNREINMGFTFTDGTFLLMKGY 65
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 EVVVVGAGLAGLAALHLLGAGRRVTVVEREDVPGGRAGLLESGGFRIDTGTPTVLTMPDL 69
QY 66 LDEMFCERSESDYLFPLSPMYRLLY-DDRDIFVYSDRENMRARLQRFDECTDGY 123
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
70 VEDAFAAVGRMADRLLEIRLAPAYRPAFGSQDLVDHDTGAAMEAAVEEFAGARQAVGY 129
QY 124 -----EQFMEOERKRF-----NALYPCITRDYSSLSKFLSLDIKALPWLAPPK 167
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 RLRLWLRLYRVQMRFFIDTNFDSPLQLAHPDLAR-LAALGGFGRDLA----- 177
QY 168 SVFNILQOYFNQEKRLAFQFSKYLGMSPWCEPALFTMLPYLEHEYGIYHVKGGLNRIRIA 227
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 ----RIGHVSDLELRVFSFOALYAGVPPARALAAAYAVIAYMDTVAGVYFPPGGHEALP 233
QY 228 AAMQAVIAENGGEIHLNSELIESELIENGAAGKVKLOHGAELRGDEV-IINADFAHAMTHL 286
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 RAMADAASDAGAVLYRQVTRL-----ERSGDRVTAVTVDQEHIPCD 277
QY 287 V--KPGVLKYY-----TPENLKOREYSCFTFMYLGLDKIY-DLPHTTVFAKDYTNIR 338
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 VVLTPTDLPVSYRLILGRTPRPLRHSPPSAVILHTGTRTPDPLAHTTISFGAANKTEH 337
QY 339 NIFDNKTLTDFFSYVQNASDSDSLAPACKSALYLVMPNNDG---LDWAHCQNV 395
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
338 ELFTGTGRMSDPSLLITRPTATDPSLAPGKHLYVLAPCPNTEVPGVGRNRELQPRYR 397
QY 396 EQVLDTLGARLGLSDIRAHIECEKIITPOTWETDSHVYKGTATFSLSHKFSQMLYWRPHNR 455
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
398 DELLAEI-ERREMPGLGAIEEBGLVTPVDWTAQGEA-ACTPPSVAHTTPTQTGPFPRNL 455
QY 456 FEELANCYLVGGGTHPGSLPTIVESARISAKLI 490
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 VRGTVNAVLAGCGTTPGCVGVTPLISGLKLAERIT 490

```

```

RESULT 12
CRTI_STRSE STANDARD; PRT; 508 AA.
AC P54971;
AC 01-OCT-1996 (Rel. 34, Created)
AC 01-OCT-1996 (Rel. 34, Last sequence update)
AC 28-FEB-2003 (Rel. 41, Last annotation update)
AC Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
AC CRTI.
AC Streptomyces setonii.
AC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
AC Streptomycineae; Streptomycetaceae; Streptomyces.
AC NCBI_TaxID=38315;
AC [1]
AC SEQUENCE FROM N.A.
AC STRAIN=ISP 5395;
AC Hoshi K.;
AC Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
AC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
AC intermediary of phytofluene by the symmetrical introduction of two
AC double bonds at the C-11 and C-11' positions of phytoene.

```

-1- COFACTOR: FAD (Probable).
 -1- PATHWAY: Carotenoid biosynthesis.
 -1- SIMILARITY: Belongs to the phytoene dehydrogenase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL: D55723; BAA09537.1; --
 InterPro: IPR000739; Adrnxr_reductase.
 InterPro: IPR002937; Amino oxidase.
 InterPro: IPR008150; Bac_phytoene_ch.
 InterPro: IPR000205; NAD_BS.
 InterPro: IPR008151; Phycn_dehydro.
 Pfam: PF01593; Amino oxidase; 1.
 PRINTS: PR00419; ADXRDTASE.
 ProDom: PD139017; Phycn dehydro; 1.
 ProSITE: PS00982; PHYTOENE_DH; 1.
 Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
 T SEQUENCE 508 AA; 54610 MW; 4B6DEF0C076D51CB5 CRC64;
 Query Match 18.2%; Score 493; DB 1; Length 508;
 Best Local Similarity 27.2%; Pred. No. 1.8e-29;
 Matches 138; Conservative 98; Mismatches 233; Indels 48; Gaps 12;
 Y 6 HIIIVAGGGGCGGAGMLLSQSGKYSIFDKHAEIGENRPNMNGFTPTGPTFLMKGV 65
 b 10 HVVVGAGLAGLAALHLLGAGRSVTWVEQVPGCRAGLLETDGFRVDTGPTVLTMPDL 69
 Y 66 LDMPELCERRSEDEYLFPLSPMYRLY-DDRDFVYSDRNWEAEQRFVDEGDTGY 123
 b 70 VEEAFVAPCEPADRLRLRLPATRAPADCSQGVDTGGAERAVEQFAGARQAVGY 129
 Y 124 -----EQFMQEQRKR-----FNALYPCITRDYSSLSKPSFLSLDLKALPWLAPPK 167
 b 130 RLRLIMLERLYRVQMERFIDANFDSFQVLPVHPDLAR-LAALGGFGLDA----- 177
 Y 168 SVFNILGVFNQKRLAFCFSKYLGMSPWECPALFTMLPYLEHYGIIYVVGGLNRIA 227
 b 178 ----RIGHEVSDERLRRVFSFOLYAGVPPARALAAVIAVMOTVAGVYPPFGGHALP 233
 Y 228 AAMAQVIAENGCEIHLNSESIEBLIENGAAKGVKLGHAELRGDEVIINADFAHAMTHLV 287
 b 234 RAMADAAADAGASFRYQSQTVEL-RESGDRVTAVTDOERTACCAVLTPLPVSYRLLG 292
 Y 288 KGVLYKKTYPENLKOREYSCSTFMVLGLDKIY-DLPHHTVFAKDYTNIRNFDKNTL 346
 b 293 R-----SPHRPLPLRHSPPSAVILHAGTDRTPNLAHHTISFGAAWKSTPHELTRTCEL 345
 Y 347 TDDFSFYQVNASASDLSAPACKSALYLVVPMNDSG--LDWCAHQNVREKQVLDLIG 403
 b 346 MSDPSLLITRTPASDPSLAPSKHLHYVLACPNTEVPGVREHRELGPVYRDELLAEL- 404
 Y 404 ALGLSDIRAHIECKIITPQWETDEHYKGAFTPSLSHKFSQMLYRPHRPFELANCY 463
 b 405 ERREMPGLGSALEVEGLVTPVDWTAQCHA-AGTSPVSAHTPPQTGPFRPGNLVRGTNAV 463
 Y 464 LVGGGTHPGSGLETPYESARISAKLIS 490
 b 464 LAGCGTTPGVGVFTVLISGLAAQRIT 490
 RESULT 13
 YTD RHOSH
 CRTD RHOSH STANDARD; PRT; 495 AA.
 Q01671
 01-JUL-1993 (Rel. 26, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methoxyneurosporene dehydrogenase (EC 1.14.99.-).
 GN CRTD.
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 OX NCBI_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RX MEDLINE=92307398; PubMed=1612412;
 RA Gari E., Toledo J.C., Gibert I., Barbe J.;
 RT "Nucleotide sequence of the methoxyneurosporene dehydrogenase gene
 from Rhodobacter sphaeroides: comparison with other bacterial
 carotenoid dehydrogenases."; FEMS Microbiol. Lett. 72:103-108 (1992).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RX MEDLINE=95238278; PubMed=7721699;
 RA Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;
 RT "Complete DNA sequence, specific TMS insertion map, and gene
 assignment of the carotenoid biosynthesis pathway of Rhodobacter
 sphaeroides."; J. Bacteriol. 177:2064-2073 (1995).
 RL J. Bacteriol. 177:2064-2073 (1995).
 CC -1- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDE NE OR
 CC METHOXNEUROSPORENE TO SPHEROIDE NE.
 CC -1- COFACTOR: FAD (Probable).
 CC -1- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 CC -1- SIMILARITY: Belongs to the phytoene dehydrogenase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X63204; CAA44886.1; --
 DR EMBL: AJ010302; CAB38743.1; ALT_SEQ.
 DR PIR: S23633; S23633.
 DR InterPro: IPR002937; Amino oxidase.
 DR InterPro: IPR008150; Bac_phytoene_ch.
 DR InterPro: IPR000205; NAD_BS.
 DR Pfam: PF01593; Amino oxidase; 1.
 DR PROSITE: PS00982; PHYTOENE_DH; 1.
 DR Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 DR Oxidoreductase; FAD; Flavoprotein; NAD.
 DR NP_BIND 9 42 FAD (ADP PART) (POTENTIAL).
 FT CONFLICT 115 115 A -> G (IN REF. 2).
 FT CONFLICT 157 157 T -> P (IN REF. 2).
 FT CONFLICT 170 170 L -> M (IN REF. 2).
 FT CONFLICT 273 273 L -> I (IN REF. 2).
 FT CONFLICT 440 443 PHGA -> ATGP (IN REF. 1).
 FT SEQUENCE 495 AA; 52900 MW; 62227931415B253E CRC64;
 SQ
 Query Match 12.0%; Score 326; DB 1; Length 495;
 Best Local Similarity 22.7%; Pred. No. 5.9e-17;
 Matches 115; Conservative 98; Mismatches 243; Indels 50; Gaps 13;
 QY 7 IIVVAGGGGCGGAGMLLSQSGKYSIFDKHAEIGENRPNMNGFTPTGPTFLMKGV 66
 Db 8 VVVVGAGLAGLAALHLLGAGRSVTWVEQVPGCRAGLLETDGFRVDTGPTVLTREVF 67
 QY 67 DEMFELCERRSEDEYLFPLSPMYRLY-DDRDFVYSDRNWEAEQRFVDEGDTGY 113
 Db 68 DDIFEVCGCKLDHHLTLPLQPLARHMLDGLDSTDLTTDLANVEAAVAPAGAREAPAFR 127
 QY 114 RVPDEGDTGYEQFMQEQRKRNALYPCITRDYSSLSKPSFLSLDLKALPWL 163
 Db 128 RPHDLRLARLYDAF---DRPMRAARPDRLRAITATGALKAPRTWPLPGLTDLRLRL--- 181

```
QY 164 APPKSVFNLLGOYFNQEKORLAPCFQSKYLQNSPWCPCALFTMLPYLEHEYGIYHVKGGL 223
DB 182 -----FFDRRLRLQFRGYATYVGGTYPGAPGVLAIT-WAAEARGVWAEIGGM 228
QY 224 NRIAAACVIAENGGEHLNSETESLIIENGAAGKGVKLOHGAELRGDEVLINADFAHAM 283
DB 229 HRLALALARLADDOGVRLRYCAPVAGILRQCGRTGTVOLADGRTLPADHIVFNGDPAALL 288
QY 284 THLVKPGVLKYYTENIKQREYSCTFWMLYGLDKIYDLPHHTIVFADYTTNIRIFDN 343
DB 289 AGCLGDGQDQAVPDRIPRSLSAWWS-YAARASGPPVLVHNVFFADDPREFFGPIAAG 347
QY 344 KTLTDDREFFYQNASADSLAPAGKSAALYVLPVPMNDSGLDQAHQCNVQVLDTLG 403
DB 348 Q-MEDAILYICADRSQGLPDGPPEFIINNGPQRPAPKEDPAC-----RSRTDRL- 402
QY 404 ARLGLSDIRAHIECKIITPOTWETDEHYKATGATFSLKFSQMLYWRPNRFEELANCY 463
DB 403 RQFGLTFDPVGE-TSLTAPSGFASLFPASQSGIYGLSPHGLASLKRPLAR-TALPGLW 460
QY 464 LVGSGTHGSGLPITYESARISAKLI 489
DB 461 LAGGAGHPGAGVPAALSGRHAARAI 486

RESULT 14
RTD_RHOCA
ID_CRTD_RHOCA STANDARD; PRT; 494 AA.
AC P17059;
YT 01-AUG-1990 (Rel. 15, Created)
YT 01-AUG-1990 (Rel. 15, Last sequence update)
YT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methoxyneprosoprene dehydrogenase [SC 1.14.99.-].
CRD.
X Rhodobacter capsulatus (Rhodospseudomonas capsulata).
X Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
X Rhodobacteraceae; Rhodobacter.
X NCBI_TaxID=1061;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SB1003 / St Louis, and BEC404;
X MEDLINE=93113693; PubMed=2747617;
RA Armstrong G.A., Alberti M., Leach F., Hearst J.E.;
RT "Nucleotide sequence, organization, and nature of the protein
RT products of the carotenoid biosynthesis gene cluster of Rhodobacter
RT capsulatus."
RL Mol. Gen. Genet. 216:254-268(1989).
RN [2]
SIMILARITY TO CAROTENOID DESATURASES.
X MEDLINE=90368827; PubMed=214293;
RA Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
RT "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
RT crassa are structurally and functionally conserved and contain
RT domains homologous to flavoprotein disulfide oxidoreductases."
RL J. Biol. Chem. 265:16020-16024(1990).
X -!- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR
X METHOXYNEUROSPORENE TO SPHEROIDENE.
X -!- COFACTOR: FAD (Probable).
X -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
X -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
X This SWISS-PROT entry is copyright. It is produced through a collaboration
X between the Swiss Institute of Bioinformatics and the EMBL outstation -
X the European Bioinformatics Institute. There are no restrictions on its
X use by non-profit institutions as long as its content is in no way
X modified and this statement is not removed. Usage by and for commercial
X entities requires a license agreement (See http://www.isb-sib.ch/announcement/
X or send an email to license@isb-sib.ch).
X EMBL; X52291; CAA36537.1; --
X EMBL; Z11165; CAA77544.1; --
X PIR; S04406; S04406.
X InterPro; IPR009937; Amino oxidase.
```

```
DR InterPro; IPR008150; Bac_phytoene dh.
DR InterPro; IPR008151; Phytan dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phytan dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 8 41 PAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 494 AA; 52312 MW; D1180A023FFB5A9 CRC64;

Query Match
Best Local Similarity 12.0%; Score 325; DB 1; Length 494;
Matches 127; Conservative 94; Mismatches 222; Indels 88; Gaps 21;

QY 1 MANTKHIIIVAGAGGGLCAGMLLSQRFKVSIFPKHAEIGGRNRPINMGFTFTPTGTPL 60
DB 1 KRSETDVVIGARMGGLAAIAAGAAAGLVTVVEAGDAPGCKARAVPTPGPADTGTPTVL 60
QY 61 LMKGVLDMPFELCRRRSEDYLEFLPLSPMYRLLY-DDRDIFVYSDRE-----NM 108
DB 61 TMRHVLDALFACCTRAEHLTLIPLRLARHFWPDGSSLDLFTDTEANTEARAFAGDK 120
QY 109 RAEIQRVDEGTG-YTQFMQEQRKRNALYP-----CITRD--YSSIKSFLSL-DL 156
DB 121 EAAAFRRPDLTTGLWEAF---HRSVIAAPKPDILWRIAATAATVTRPOLWPAIRPGLTWRDL 177
QY 157 IKALPWLAFPKSVFNNLGOYFNQEKELAPCFQSKYLQNSPWCPCALFTMLPYLEHEYGI 216
DB 178 -----LAHFQDPLQLQRFQRTATYVGGPRTPAVLISLIWQAEVQ-GV 220
QY 217 YHVKGGLNRIAAAOVIAENGSEIHLNSETESLIIENGAAGKGVKLOHGAELRGDEVLIN 276
DB 221 WAIREGNHVAALARAARVAEAKGVRFHYGAKAKRIVRKEGRTVAVEIETGVSIPOGACIFN 280
QY 277 ADFAHAMTHLVKRGVLK-----KYTPENLQREYSCSTFFMLYGLDKI-YDLPHTTI 327
DB 281 GD-----PGALRDGLLGDAARASMEKSRPAPSLSAWVAFGATPIGVDLAHNV 330
QY 328 VFAKDYTTNIRNIFDNKTLTDDREFFYQNASADSL-APAGKSAALYLV-----PM 378
DB 331 FFTADPELEFGPIGAGE-MPEEPTLYI---CAQDREMOAPVPEIERFEILMNGPAGHOPF 386
QY 379 PNNDSGLDQWQHCONVREVDLTGLARLGLSDIRAHIECKIITPOTWETDEHYKATP 438
DB 387 PQE-----EAQC---RARTFPLAA-MGLT-FSPDPETRALTPALLSRFRPFGSLGAIY 435
QY 439 SLSHKFSOMLYWRPNRPFELANLYVGGGTHPGSGLPITYESARISAKLI 489
DB 436 GGSPEGLTATFRPLAR-TGLKGLYLAGGCTHFGAGVPMALTSGTHAARAL 485

RESULT 15
P49_STRLI
ID_P49_STRLI STANDARD; PRT; 469 AA.
AC P06108;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE P49 protein.
GN P49.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=66 / 1326;
RX MEDLINE=87231086; PubMed=34531116;
RA Burnett W.V., Henner J., Eckhardt T.;
RT "The nucleotide sequence of the gene coding for XP55, a major
RT secreted protein from Streptomyces lividans."
RL Nucleic Acids Res. 15:3926-3926(1987).
CC -----
```


This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Y00142; CAA68336.1; -
PIR; S09189; S09189.
InterPro: IPR000759; Adrndx reductase.
InterPro: IPR000205; NAD_B5
InterPro: IPR008151; Phytin dehydro.
PRINTS; PR00419; ADXRDASE.
ProDom: PD139017; Phytin dehydro; 1.
SEQUENCE 469 AA; 49740 MW; C2AE7533C7C701CB CRC64;

Query Match 6.5%; Score 175; DB 1; Length 469;
Best Local Similarity 21.8%; Pred. No. 1.2e-05;
Matches 118; Conservative 70; Mismatches 221; Indels 132; Gaps 26;
8 IIVGAGPGICAGMILLSORGFKVIFDKHAEIGG--RNRPINMGFTFD-----T 55
5 VVVGAGPGLTAARVLAARVFAVFAAGTVGGGARTTELTLPGRHDPGSAHPGGIN 64
56 GPTFLMKGVLDMEFELCERRSEDVLEFL-PLSPMYRLLYDDRDIFVYSDRENRAELQR 114
65 SPAF---RGLPLERYG-----LEWLHGLPMAH--PPDGSAAVLSRSGVETAASFG 111
115 VFDEGTGDTGQFMGEQRKFNALYPCITRDYSSLSKSLSLDLIKALP-----WLA 164
112 ARDAGP--YRRUIE----RFLPRWDTLARDFMSLP-----LTALPRDPVTLARFGLVG 158
165 FPKSVFNILGOYENOEKMLAFCSQKVLGMSPP--WECPALFTMLPYLHEHYGIYHVKG 221
159 LPPSTW--LMRRFRDEKATLFAGLVAHV--MAPLGGFATGAGLVAALAHARGWPVARG 215
222 GLNRJAAAMAQVIAENGGEIHLNSESIL-----IIENGAAGVKL-----OHGAEL 268
216 GSQISDALTAYLKDLGGAVHTDYEVKRLDDLPPARAYVLDTSPTALARIAGLGSYVANY 275
269 RGEVINADFAHATHLVKGVKKYTPENLKQREYSCSTFMYLGLDKIYDLPHHTIV 328
276 RYGPSVFKIDYA-----LDGP---VPWTAEEPR-----SAGTVOIGAD----- 310
329 FAKDYTTNIRNIFDNKLTLD-----DFSGYVQNASASDDSLAPAGSKALYVLPMPNDS 383
311 -----STEIGALHAPSGTDRAPERFLITQ-PSVADPTAPAGKGVFWAYGVPNG-- 362
384 GLDQAHQONVREQVLDTLGRLGLSDIRAHIECEKIITPQTWETDEHYKATFS----- 439
363 ---WNGDLTDAZEROLERFA-----PGFEDRVLARATAGPAELARNANYVGGDISGAV 414
440 -----LSHKFSOMLYWRPHNRFEELANCYLVGCGTHPGSGLPTIYESARISAKLISOKH 493
415 SGLQLLRPKISLFFPYSTH-----PAVFICSSATPPGPGVHGM--SGHNRKAKWWRRL 466
494 R 494
467 R 467

earch completed: February 29, 2004, 14:44:59
ob time : 12.6124 secs

1 protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 14.2234 Seconds
(without alignments)
3455.835 Million cell updates/sec

File: US-09-941-947A-22

Effect score: 2711

Sequence: 1 MANTYKHIIIVAGGGLGAG.....KRVRFKDIHSAWLKKAKA 511

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Aligned: 283366 seqs, 96191526 residues

Hit number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	929.5	34.3	490	2 AH1199	phytoene dehydroge
2	915.5	33.8	488	2 AF1557	phytoene dehydroge
3	883.5	32.6	517	2 T31463	probable diapophyt
4	825.5	30.5	529	2 S35306	phytoene dehydroge
5	794.5	29.3	499	2 AG2599	zeta-carotene desa
6	794	29.3	517	2 S32169	hypothetical prote
7	783.5	28.9	499	2 S43324	zeta-carotene desa
8	743	27.4	524	2 A32617	phytoene dehydroge
9	737	27.2	518	2 S49620	phytoene dehydroge
10	735	27.1	518	2 T50745	phytoene dehydroge
11	725	26.7	511	2 T50910	phytoene dehydroge
12	713.5	26.3	492	2 S25986	phytoene dehydroge
13	704.5	26.0	492	2 A33120	phytoene dehydroge
14	702.5	25.9	502	2 B90061	squalene synthase
15	700.5	25.8	492	2 D37802	phytoene dehydroge
16	686	25.3	583	2 S43139	phytoene dehydroge
17	681	25.1	595	2 A35919	carotenoid biosynt
18	673	24.8	582	2 T46822	phytoene desaturas
19	651.5	24.0	548	2 C75466	phytoene dehydroge
20	613.5	22.6	448	2 B55548	crtn protein - Sta
21	607	22.4	621	2 T48646	phytoene dehydroge
22	596	22.0	514	2 E59108	phytoene dehydroge
23	564.5	20.8	523	2 T36968	probable phytoene
24	547.5	20.2	512	2 H84320	phytoene dehydroge
25	504	18.6	485	2 S32171	hydroxyneurosporen
26	502.5	18.5	497	2 E90061	hypothetical prote
27	485.5	17.9	536	2 B84327	phytoene dehydroge
28	472.5	17.4	528	2 T51119	phytoene desaturas
29	404	14.9	525	2 T50893	methoxyneurosporen

30	392.5	14.5	454	2 A99470	phytoene dehydroge
31	354	13.1	498	2 H83880	hypothetical prote
32	332.5	12.3	542	2 S76617	hypothetical prote
33	325	12.0	494	2 S04406	methoxyneurosporen
34	324	12.0	495	2 S23633	methoxyneurosporen
35	323	11.9	511	2 E75561	probable phytoene
36	322	11.9	486	2 S49624	methoxyneurosporen
37	314	11.6	565	2 A12273	hypothetical prote
38	290	10.7	486	2 T50749	methoxyneurosporen
39	278	10.3	506	2 T34971	probable carotenoi
40	276.5	10.2	587	2 A86203	hypothetical prote
41	273.5	10.1	503	2 AC2446	hypothetical prote
42	256.5	9.5	506	2 AB2064	hypothetical prote
43	249	9.2	518	2 G90413	phytoene dehydroge
44	249	9.2	574	2 A96612	hypothetical prote
45	248	9.1	507	2 S74689	hypothetical prote

ALIGNMENTS

RESULT 1

AH1199

phytoene dehydrogenase homolog lmo1000 [imported] - Listeria monocytogenes (strain EGD-e
C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

A:Accession: AH1199

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1199

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-490 <GLA>

A:Cross-references: GB:NC 003210; PIDN:CAC99078.1; PID:G16410402; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo1000

Query Match 34.3%; Score 929.5; DB 2; Length 490;

Best Local Similarity 37.6%; Pred. No. 4.3e-62;

Matches 184; Conservative 98; Mismatches 199; Indels 9; Gaps 1;

QY	5	KHIIIVGAGPGGLCAGMLLSORGFKVSIFDKHAIIGGNNRPINNNGTETDGTFLMKG	64
DB	7	KKAIIVGAGPGGLCAGMLLSOLGQVQVIYKNDRIKGRGRTALHRMGKYSFDVGPALNTW	66
QY	65	VLDENFELCERRSDYLFELPLSPMYRLLYDDRDIFVYSDENNRAELQRFDEGTGCE	124
DB	67	VLTSLFMDCNELIDYVSLPILNHTLYPKDITFLYSQDSETKAVIQTFPGESDGF	126
QY	125	QPMQERKFNALYPCITRDYSSLSKSLDLIKALPWLAPPKSVFNNLQYFQNERML	184
DB	127	RFMKENTKMLYISPLNQFNYSLSDFRPTTLEAIPLSLTLGRSLMDLGRYFNKYLRL	186
QY	185	APCFOSKVLGSPWECAPALFTMLPVLHEVGIYHVKGLENRAAAMAQVLAENGGEIHLN	244
DB	187	AFSLQVRILGSLPMDIPAYSIIPFSEVYGTFTPIGQNKIVAMQVVTENKGFEN	246
QY	245	SEIBSLIIENGAAGKVLQHGAEGLRGEVIINADPAHAMTHLVKPGVLKKTYPENLKRE	304
DB	247	SEVTEFESNGEITGAVLANGKTIEADYFTNLDIFYSLTN-----EHPDKLTKE	297
QY	305	YSCSPFMYLTGLDKTYDLPHHTIVFAXDYTTNIRNIFDNKTLTDDFSFYQNASASDLS	364
DB	298	YSSAFIYLTGLKTVLPFPHSHOSIIPENYREFAINTWHKKILSKDIAHLNTPSATNTM	357
QY	365	APAGKSALYVLPMPNNDSGLDWOAHQCONVREQVLDLTGLARGLSLIRAHIECEKII	424

```
Db 358 APINSHSIRIMVPEINTSNDWKEKTPAPQOLVETVGRSLIPDLSEIIEEYIITPI 417
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 425 TWETDEHYKATFSLGKTSQMLYWRPHNRFELANCYLVGCGTHPGSLPTIYESARI 484
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 418 DWENKYHWHGCAIFGLQHLWQHGYLHPSKSKPKFKNLFIYGAGAMSGSLPLIENAQI 477
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Qy 485 SAKLISQKHR 494
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 478 ATQKFLQKEX 487
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 2
AF1557
phytoene dehydrogenase homolog lin0999 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1557
C:Reference: Dominguez-Bernal, G.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
      J.; Glaser, P.; Frangeul, L.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Pshi, H.
      Science 294, 849-852, 2001
C:Author: Kretz, J.; Kunz, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
      x, C.; Schluerer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
      J.; Title: Comparative genomics of Listeria species.
C:Reference number: AB1077; MUID:21537279; PMID:11679669
C:Accession: AF1557
C>Status: preliminary
C:Molecule type: DNA
C:Residues: 1-488 <GLA>
C:Cross-references: GB:AL592022; PIDN:CAC96230.1; PID:G16413458; GSPDB:GN00178
C:Experimental source: strain Clip1262
C:Genetics:
C:Gene: lin0999

Query Match 33.8%; Score 915.5; DB 2; Length 488;
Best Local Similarity 37.4%; Pred. No. 4.9e-61;
Matches 185; Conservative 95; Mismatches 205; Indels 9; Gaps 2;

Qy 1 MANTKHHIIVGAGPGGLCAGMLLSQSGFKVSIIDKHAIEIGRNRPINMGFTEDTGTPL 60
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 1 MESKKTAIAGGPGGLATCMLLSQGYQVSIYEKNDRIIGRTALHMGKYSDVQPSAL 60
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 61 LMKGVLDMEFELCRRSEDEYLFPLSPMYRLLYDDRDIFVYSDRENMRABLQVDFEGT 120
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 61 TWITVLTSLFDCNRNILDYVSLPPIPIHTLVFKOITPPLSYDSETKVIKTYPGEE 120
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 121 DGEQFMEOERKRFNALYPCITRDYSSLSKFLSLDLIKALPWLAPPKSVFNNLQYFNQ 180
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 121 DGFPRFMENTKMLYISPLNQFNYSLLPFFRPTTIRALPSLTGRSLMDDLRATFNSK 180
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 181 XRLAFQCSKYLQMSPECPALFTMLPYLEHYGIYHVKGGLNRIAAAMAQVIAENG 240
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 181 YLRIFAFLQKRYLQMSPDIPAAVSIIPFSEYVPGTFRPHIGQGNKIYEAQQVVMENK 240
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 241 IHLNSEIESLIENGAKGVKLOHGAELGDEVIINADFAHAMTHLVKPGVLKKTYPEN 300
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 241 FYFNSEVTSPTNKEKTSATLANGQIEADYFTNLDFIYLSUTSNK---VKE----- 291
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 301 KQREYSCSTFMYLGLDKIYDLPHHTIIVPAKDYTTNIRNIFDNKLTLDGDFSVYQNAS 360
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 292 KEKEYSSAFMYLGLATVLPFSSHQSIIIPDNYREFANNTIHKKILSKDLAIIHNP 351
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 361 DDSDLPAGKSNLYLVPMPNNDGLDQWQAHQCNVREQLVDTLGLARGLSDIRAHIECE 420
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 352 DNTWAPMNHSSVRIMVAPNTTSGIDWKEQTAPFRLQILDIVKARLEIPDELYIEEYI 411
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 421 ITPQTWETDEHYKATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHPGSGGLPTI 480
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 412 ITPDWEKHYHVYGAIFGLQHLWHQOGLFEPKAKSPKFKNLVVGAGAMSSSLPPIIE 471
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 481 SARISAKLISQKHR 494
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Db 472 NAQIATKFLQKEX 485
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 3
T31463
probable diaphytoene dehydrogenase crtn - Helicobacillus mobilis
C:Species: Helicobacillus mobilis
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 15-Sep-2000
C:Accession: T31463
C:Reference: RiXiong, J.; Inoue, K.; Bauer, C.E.
      Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998
C:Title: Tracking molecular evolution of photosynthesis by characterization of a major
      A:Reference number: 221036; MUID:99061957; PMID:9843979
C:Accession: T31463
C>Status: preliminary; translated from GB/EMBL/DBJ
C:Molecule type: DNA
C:Residues: 1-517 <XIO>
C:Cross-references: EMBL:AF080002; NID:G3820536; PID:G3820561; PIDN:AAC84034.1
C:Genetics:
C:Gene: crtn
C:Superfamily: phytoene dehydrogenase

Query Match 32.6%; Score 883.5; DB 2; Length 517;
Best Local Similarity 35.2%; Pred. No. 1.4e-58;
Matches 174; Conservative 117; Mismatches 194; Indels 9; Gaps 6;

Qy 3 NTKHHIIVGAGPGGLCAGMLLSQSGFKVSIIDKHAIEIGRNRPINMGFTEDTGTPL 62
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 7 NLGKVIIVGSGAGGSAARLANQGDVTVLEKATPGGRLSAIQAGYSIDVGTPIWM 66
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 63 KGVLDMEFELCRRSEDEYLFPLSPMYRLLYDDRDIFVYS-DRENMRABLQVDFEGT 121
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 67 NDVFHQVFKDLGRNIEDYDLVRVPCYHLHFTDGTCKMPSIDLKELLDEIRSFNPD 126
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 122 GYEQFMEOERKRFN-ALYPCITRDYSSLSKFLSLDLIKALPWLAPPKSVFNNLQYFN 180
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 127 GYLRYLAQIHHRYQVAREKFEKSTFKPSDFNIDTLIGMMQRLTNMNYDDIARFIK 186
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 181 XRLAFQCSKYLQMSPECPALFTMLPYLEHY-GIYHVKGGLNRIAAAMAQVIAENG 239
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 187 RLRLSLFQALYLCVSPDAPSIYTLGYVEHGLSGVWYKPGGNVATQALVLLGEFG 246
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 240 EHLNSEIESLIENGAKGVKLOHGAELGDEVIINADFAHAMTHLVKPGVLKKTYPEN 299
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 247 SLRYNAEVEQLIEQGRAGVGRVLAKEVLSADPFTYMNVLVPASHRGKTYPOK 306
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 300 LKQREYSCSTFMYLGLDKIYDLPH-HTIIVPAKDYTTNIRNIFDNKLTLDGDFSVYQ 358
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 307 LNSKEVTCGAPMLYLVGNRRYDNLHLENIYFTDYKYSMDLFTROQLPQDPAMYV 366
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 359 ASDSDLPAGKSNLYLVPMPNNDGLDQWQAHQCNVREQLVDTLGLARGLSDIRAHIE 418
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 367 KYDSDVAPPGKDIYLVLPVVPVNLSSGIDMKKETHRYRELVIKKL-BEQGVTDLSKH 425
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 419 KIITPQTWETDEHYKATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHPGSGGLPT 478
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 426 RIYTPETQNFNTYQGAFLAPSLFQSGYFRPHIKSKEVPNLYFGSGVHPGGVPVV 485
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Qy 479 YESARISAKLISQK 492
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 486 L-----VCGKLVSQ 495
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 4
S35306
phytoene dehydrogenase (EC 1.3.-.-) - Myxococcus xanthus
C:Alternate names: phytoene desaturase
C:Species: Myxococcus xanthus
C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 29-Sep-1999
C:Accession: S35306; S27594
C:Reference: R; Ruiz-Vazquez, R.; Murillo, F.J.
      EMBO J. 12, 1265-1275, 1993
C:Title: Growth phase dependence of the activation of a bacterial gene for carotenoid s;
```

Reference number: S35306; MUID:93223667; PMID:8467787

Accession: S35306

Molecule type: DNA

Residues: 1-529 <FON>

Cross-References: EMBL:M94727; NID:g150079; PID:AAA25390.1; PID:g150080

Genetics:

Gene: carC

Superfamily: phytoene dehydrogenase

Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 30.5%; Score 825.5; DB 2; Length 529;
Best Local Similarity 35.1%; Pred. No. 3.3e-54;
Matches 179; Conservative 106; Mismatches 192; Indels 33; Gaps 10;

3 NTKHIIIVGAGCGGLCAGMLLSQSGFKVSIQFKHABIGGRNRPINMNG---FTFDGPTF 59

4 SKKVAIVGAGCGGLTATRLAGLQVVEIPEAABRVGGRMRGPEVDVSYAFDTGTILQLP 61

5 SVRHVIVGAGCGGLSAINALAGOGFRTVVEKDAVFGGKGLTIGASGEYAVDTGPSI 66

60 LLMGVLDMEFELCERSEDYLFPLSPMYRL-LYDDRDI FVYSDRENRAELQRFVDE 118

67 LQLPGVLEQIFRAARLEEDYKLLPLDVNTRVHFWDGTHTLDTTRHLDRMEALEAKFGPR 126

119 GTDGEVPEQERKFNALYF---CITRDYSSLSKLSLSDLIKALPWLAPF-----KS 168

127 CASALRQWEDGKRYGAYOKFICTSAD-----NLGYAPWRLAPTLPKFWOT 176

169 VFNLLGQYFNOEKRLAFCSQSKYLGMSPWECALFTMLPYLEHEYGIYHVKGGLNRIAA 228

177 LYRQLDGFHDDRYALAYSPVIFGLHPTTCSSVPFVIFPELAFGWVHEGGERLSR 236

229 AMAQVIAENGSEIHNLSEIESLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLV- 287

237 GMMRCARDLGATFMGTGPEVKKRVDRAGVGVLVGGEVLDADAVVWADLAYAARSLIP 296

288 ---KPGVLKXKTPENLKOREYSCSTFMLYGLDKIY-DLPHHTIVFAKDYTTNIRNIFON 343

297 AEAREG---SRUTDAALERAKYSCSTFMAYYGLDITVYADLPHLLYLSSESARRTDRALED 354

344 KTL-TDDPSFVQNASDSDSLAPAGKSALYVLVPMNDSGLDQWQAHQCNVREQLDITL 402

355 RHVLDLEDPPFYVCPGVTDPSGAPAGHSTLVLTPTNGRPVDMVKTQALRERIPAML 414

403 GARLGLSDIRAHIECEKIIPTQWETDEHYVKGATFSLSHKFSOMLYWRPHNPFELANC 462

415 -EKVLGKVRREHREERYFTAEWRDDFNVRGAVPNLSHTWLGRLPKVKNRDIEGL 473

463 YLVGGGTHPGSLPTIYESARISAKLSQK 492

474 YFVGGGTHPGSLPTIYESANIAADYLTR 503

SULT 5

42509

ta-carotene desaturase [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha

Species: Nostoc sp. PCC 7120

Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

Accession: AG2509

Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Y.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

IA Res. 8, 205-213, 2001

Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

Reference number: AB1807; MUID:21595285; PMID:11759840

Accession: AG2509

Status: preliminary

Molecule type: DNA

Residues: 1-499 <KUR>

Cross-References: GB:BA000020; PID:BA878339.1; PID:g17135793; GSPDB:GN00180

Experimental source: strain PCC 7120

Genetics:

Gene: ali7255

Genome: plasmid

Superfamily: phytoene dehydrogenase

Query Match 29.3%; Score 794; DB 2; Length 517;

Best Local Similarity 35.8%; Pred. No. 7.4e-52;

Matches 177; Conservative 106; Mismatches 201; Indels 10; Gaps 8;

5 KHIIVGAGCGGLCAGMLLSQSGFKVSIQFKHABIGGRNRPINMNGFTFDGPTFLMKG 64

8 RRIIVVGAGVGLAAARLAHQGDVQVFEKTCQGGRCNQLQVDGFTWIDGFTIVLMPE 67

Query Match 29.3%; Score 794.5; DB 2; Length 499;

Best Local Similarity 35.4%; Pred. No. 6.4e-52;

Matches 177; Conservative 102; Mismatches 200; Indels 21; Gaps 10;

4 TKHIIIVGAGCGGLCAGMLLSQSGFKVSIQFKHABIGGRNRPINMNGFTFDGPTFLMKX 63

2 SKKVAIVGAGCGGLTATRLAGLQVVEIPEAABRVGGRMRGPEVDVSYAFDTGTILQLP 61

64 GVLDEMELCERSEDYLFPLSPMYRL-LYDDRDI FVYSDRENRAELQRFVDEGTDG 122

62 HLYKELFEEAGLNADVQLKRLPEYTRLPKFDWTDQTSQSKTQLATLRSDLPLA 121

123 YEQFMEOQERKFNALY-PCITRDYSSLSKLSLSDLIKAL---FWLAPPKSVFNNLQYF 177

122 FDRWYSEHIRKVELGYKPYLAGPARSIFGYLRPDELKMLKLSFRPW---ENLYOHFWRF 177

178 NOEKRLAFCQSKYLGMSPWECALFTMLPYLEHEYGIYHVKGGLNRIAAQAQVIAEN 237

178 QDRLVYALSYPKSGHPTVASSVFLSFLPPELFSGQVWHPVGGPRALAQGLANAQDL 237

238 GGEIHLNSEEIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYKTYT 297

238 GVKIHLSPVHQIWEQGVGELADASRHPDVTVINADFAVAVRHLLPPTSARGTYD 297

298 ENLKOREYSCSTFMLYGLDKIY-DLPHHTIVFAKDYTTNIRNI---FUNKTLTD-DPS 351

298 NKLQGMQPCSTFMLYGLDKIRRYEDLPHHGIYL---SDNIRLERPWDDSDALDETPP 353

352 FVQNASDSDSLAPAGKSALYVLVPMNDSGLDQWQAHQCNVREQLDITLGAELGLSDI 411

354 FYVCNPTIIDSNAPAGHSTLVLPVPIPTSYAVDMDIKQKSYTDFILKRLHL-LGYHNI 412

412 RAHIECEKIIPTQWETDEHYVKGATFSLSHKFSOMLYWRPHNPFELANCYLVGQTHP 471

413 EQHIVTQSCYTAQSWLDYRVHLGAVFNLSHNLTLQGFPRPPPIRSENIAGLYWTGGVHP 472

472 GSGLPITVESARISAKLSQ 491

473 GSGLLTLEASRGAAGFIHQ 492

RESULT 6

S32169

hypothetical protein 2 - Myxococcus xanthus

C:Species: Myxococcus xanthus

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 29-Sep-1999

C:Accession: S32169; S67952

R:Botella, J.; Murillo, F.; Ruiz-vazquez, R.

submitted to the EMBL Data Library, March 1993

A:Description: Nucleotide and deduced protein sequences of a carotenoid gene cluster in

A:Reference number: S32168

A:Accession: S32169

A:Molecule type: DNA

A:Residues: 1-517 <BOT>

A:Cross-References: EMBL:221955; NID:9577589; PID:CAA79956.1; PID:g288221

A:Experimental source: strain DK1050

R:Botella, J.A.; Murillo, F.J.; Ruiz-Vazquez, R.

Eur. J. Biochem. 233, 238-248, 1995

A:Title: A cluster of structural and regulatory genes for light-induced carotenogenesis

A:Reference number: S67950; MUID:96061955; PMID:7588751

A:Accession: S67952

A:Molecule type: DNA

A:Residues: 9-37;462-488 <BOW>

C:Superfamily: phytoene dehydrogenase

```

65 VLDEMFELCRRSEDIYELPLSPMYRLLYDDR-DIFVYSDRENMAELQRFVDSGTGY 123
66 VFEETFRVGRRTEDYTLRLCDPNRYKRVFRDRSDVTFTELSCAMGERLERVEDPSYARY 127
124 EQMEQERKEF-NALYPCITRDYSSLSKLSFLSLDLKALPWLAPFKSVFNNGLYQYFNQRM 182
128 LAFACGRVQYRTSLDHLVGNVYAGLRDYLSPVLARIFQVRAHRMYADVSRFFQDERL 187
183 RLAFCTQSKYLGMSPECPALFTMLPYLHEHYGIYHVKGGLNRIAAAMAQVIAEN 242
188 RAAMTFQTVLGVSPSPASPAVYGLPFTBLGVGIVFPKGLYALPOALERARREGVRFH 247
243 LNSEISLIIENGAAGVKLOHGAEIRGDEVIINADFAHAMTHLVKPGVLKYYTLENLKQ 302
248 YGAPVERILTDGGRTRGVRLGEGVEVEADAVLCNADLPYAVEKLLDP---KATTLKRKEK 304
303 REYSCSTFMVLYGLDKIY-DLPHHTIVFAKDYTNIRNFNKLTLTDDSPFYVQNASAD 361
305 LRYTSSGYMLYLGKRYPELHNNVFGDYKGSFDDIFEFPR-VPEDEPSFYVNAFTRTD 363
362 DSLAPAGKSALYVLPMPNNDGLDWAHCQNVREQVLTGLARGLSLDIRAHISCE-KI 420
364 ASLAPGKDALYVLPVPHQHPDLDNKVEGPKYRAKEFARM-AELGFPLESDDIEVERES 422
421 ITPQWTDEHYVKGATFSLSHKFSQMLYRPNRFEELANCYLVGCGTHPGSGLPTIYE 480
423 STPDMAGTENLARGSGFGLSONFTQIGPPRPSNODARVKNLFFVGASTOPGGLPTVLI 482
481 SAR-ISAKLISQKH 493
483 SARLVTERLTMWAH 496

RESULT 7
343324
zeta-carotene desaturase - Anabaena sp. (strain PCC 7120)
;Species: Anabaena sp.
;Variety: PCC 7120
;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
;Accession: S43324; S62214
;Linden, H.; Misawa, N.; Saito, T.; Sandmann, G.
Plant Mol. Biol. 24, 369-379, 1994
;Title: A novel carotenoid biosynthesis gene coding for zeta-carotene desaturase: function
;Reference number: S43324; MUID:94154256; PMID:8111038
;Accession: S43324
;Molecule type: DNA
;Residues: 1-499 <LIN>
;Cross-references: GB:D26095; NID:G439477; PIDN:BAA05091.1; PID:G439478
;Experimental source: PCC7120
;Albrecht, M.; Linden, H.; Sandmann, G.
Eur. J. Biochem. 236, 115-120, 1996
;Title: Biochemical characterization of purified zeta-carotene desaturase from Anabaena
;Reference number: S62214; MUID:96184887; PMID:8617254
;Accession: S62214
;Molecule type: protein
;Residues: 2-11 <ALB>
;Superfamily: phytoene dehydrogenase
;Keywords: carotenoid biosynthesis; membrane bound

Query Match 28.9%; Score 783.5; DB 2; Length 499;
Best Local Similarity 35.0%; Pred. No. 4.3e-51;
Matches 175; Conservative 103; Mismatches 201; Indels 21; Gaps 10;

2Y 4 TKHIIIVAGPGGICAGMLLSQRFKYSIFDKHAEIGGRNRPINMGFTPDGTPTFLMK 63
DB 2 SKKVAIVAGPGGATAIRLACLGQVEIFEAEERVGGRNGEVDYSYAFDTGTLQLP 61
2Y 64 GVLDMEFELCRRSEDIYELPLSPMYRL-LYDDRDIYVYSDRENMAELQRFVDSGTG 122
DB 62 HLYKELFEAGLNADYVQLKRLSPYTLKAFWDGTQDITSDLSQSKTQTLATLSOLPLA 121
2Y 123 YEQFMEQERKEFNALY-PCITRDYSSLSKLSFLSL-DLIKAL---PWLAPFKSVFNNGLYQYF 177

```

```

122 FDRWYSEHIRKVELGKPYLAGPARSIFCYLRPDDLMKZLSFRPW-----ENLYQHFWFF 177
178 NQEKMLAFCFQSKYLGMSPECPALFTMLPYLHEHYGIYHVKGGLNRIAAAMAQVIAEN 237
178 QDERLVYDLRPSKYLGMPHTVASSVFSLLPERSQGVVHPVGGFRALAAQGLANAQDL 237
238 GGEIHLNRIESLIIENGAAGVKLOHGAEIRGDEVIINADFAHAMTHLVKPGVLKYYT 297
238 GVKIHLSPVHQIWIDQGVRGLELADASRHQFDTVWINADFAVAVRHLLPTTSARGRYTD 297
298 ENLKREYSCSTFMVLYGLDKIY-DLPHHTIVFAKDYTNIRNI-----FDNKTLTD-DPS 351
298 NLGQWQFSCSTFMVLYGINRRYEDLPHEQIYI-----SDNIRELERPWVDDSDALDETDP 353
352 FYVQNASASDLSLAPAGKSALYVLPMPNNDGLDWAHCQNVREQVLTGLARGLSLDI 411
354 FYVCPNPIIDPSNAPAGHSTLFLVLEIPNTSAVOWMDIKQKSYTDIFLKRHL-LGYENI 412
412 RAHIECEKIITPQWTDEHYVKGATFSLSHKFSQMLYRPNRFEELANCYLVGCGTHP 471
413 EQHIVTQSCYTAQSWLDDYRVHUGAVFNLSHNLTLQGLPPRPPRPSRSENIALGLYWGAVHP 472
472 GSGLPTIYESARISAKLISQ 491
473 GSGLLTILASRSNAGFIHQ 492

```

RESULT 8

```

A32617
phytoene dehydrogenase (EC 1.3.-.-) - Rhodobacter capsulatus
N:Alternate names: phytoene desaturase
C:Species: Rhodobacter capsulatus
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 29-Sep-1999
C:Accession: A32617; S04402
R:Bartley, G.E.; Scolnik, P.A.
J. Biol. Chem. 264, 13109-13113, 1989
A:Title: Carotenoid biosynthesis in photosynthetic bacteria. Genetic characterization of
A:Reference number: A32617; MUID:89327279; PMID:2546948
A:Accession: A32617
A:Molecule type: DNA
A:Residues: 1-524 <BAR>
A:Cross-references: GB:J04969; NID:G340554; PIDN:AAA50313.1; PID:G556397
R:Armstrong, G.A.; Alberti, M.; Leach, F.; Hearst, J.E.
Mol. Gen. Genet. 216, 254-268, 1989
A:Title: Nucleotide sequence, organization, and nature of the protein products of the c.
A:Reference number: S04401; MUID:89313663; PMID:2747617
A:Accession: S04402
A:Molecule type: DNA
A:Residues: 1-524 <ARM>
A:Cross-references: EMBL:X52291; NID:G45996; PIDN:CAA36533.1; PID:G45998
A:Note: translation of codons 1-33 is not given
A:Note: the authors translated the codon GTG for residue 34 as Met
C:Genetics:
A:Gene: crtI
C:Superfamily: phytoene dehydrogenase
C:Keywords: carotenoid biosynthesis; oxidoreductase

```

```

Query Match 27.4%; Score 743; DB 2; Length 524;
Best Local Similarity 32.8%; Pred. No. 5.2e-48;
Matches 162; Conservative 103; Mismatches 211; Indels 18; Gaps 9;

```

```

QY 8 IIVAGPGGICAGMLLSQRFKYSIFDKHAEIGGRNRPINMGFTPDGTPTFLMKGVLD 67
DB 12 VVIAGLGGAAARLGAQYKVTVDRLDRPGGRSSITKGGHFDLGLPTIYVDPRLR 71
QY 68 EMFELCRRSEDIYELPLSPMYRLLYDDRDIY-VYSDRENMAELQRFVDSGTGYEQF 126
DB 72 ELNADCGEDFDKQVSLVPMEFFITIDPDGCKYATYGDGAKVAEVARISPGDVEGFRHF 131
QY 127 MEQERKEFNALYPCITRDYSSLSKLSFLSL-DLIKALP---WLAPFKSVFNNGLYQYFNQEK 182
DB 132 MWDARAYEFGYENLGR-----KFMKSLWDLIKVLPFTFGLRADRSVYGHAKMKWDDHL 186

```


Figure 1 shows a Western blot analysis of protein expression in C2C12 myotubes. The blot displays bands for p38, p38 phosphorylated, p42, p42 phosphorylated, p44, p44 phosphorylated, and GAPDH. The lanes are labeled: Control, 10^{-6} M, 10^{-5} M, 10^{-4} M, 10^{-3} M, 10^{-2} M, 10^{-1} M, and 10^0 M. Molecular weight markers (66, 43, 30, 20, 15, 10, 7.5, 5, 3 kDa) are indicated on the left. The p38 and p42 bands show increasing intensity with increasing concentration of the treatment, while the p38 phosphorylated and p42 phosphorylated bands show a decrease in intensity. GAPDH serves as a loading control and shows consistent intensity across all lanes.

!RESULT 11

```

>50910
>Phytoene dehydrogenase [imported] - Rubrivivax gelatinosus
>Species: Rubrivivax gelatinosus
>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
>Accession: T50910
>Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
>Submitted to the EMBL Data Library, November 1999
>Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosyn
>Reference number: Z52570
>Accession: T50910
>Status: preliminary; translated from GB/EXBL/DBSJ
>Molecule type: DNA
>Residues: 1-511 <NAG>
>Cross-references: EMBL:AB034704; PIDN:BAR94063.1
>Experimental source: strain IL144
>Genetics:
>Gene: crtI
>Superfamily: phytoene dehydrogenase

```

Query Match	26.7%	Score	725;	DB	2;	Length	511;
Best Local Similarity	32.4%;	Pred. No.	1.1e-46;				
Matches	158;	Conservative					
8	IIVGAGGGGLCAGMLLSQRFKYSIIPDKHAEIGERNRPNMNGFTEDTGTPLLMKGVL	67					
22	LIVSGSGGMAAAVRLAAKGRVTVLEKLDAPGGRAYVIRREGHVDEAGTTVTPYFLD	81					
68	EMFELCERRSEDYLEFLSPMYRLLYDDRDIFVYS--DRENMAELQRFDEGTDGYBQF	126					
82	ELWALAKRKTSDDIELKSLDPFYRIIRFDGCHFDYSGDPAEMAEVRRISPDAEGPERF	141					
127	MEQERKRFNALYPCITRDSYSL--KPSLSL-DLIKALP-----WLAEPKSVFNALGOY	176					
142	MREADQ-----CYELGFTTLDGKAFDTVGDIIKAAPLIVKRGW----RSLHQWVSSH	190					
177	FNQEKWRLAFCQFSKYLKMGSPWECALFTWLPVLEHYGIYHKGLNRIANAQAIVAE	236					
191	LKGPKLRITAMLSQELLIGGPFPSVTSMYALVNALERQGWVHWANGTGBELRGLVDVPEG	250					
237	NGGSEIHLNSETIESLIIENGAAKGVKLQHGAEALRGDEVIINADFAHAMTHLVKPGVLKKT	296					
251	MGGMELKAEVKRIEIVGVNGVATGVTLAGDERIPADIVVCNGDTGYLYKNLVDARWKSWT	310					
297	PENIKQREYSCTTMYLGLDKIV-DLPHHTIYPAKDYTTNIRNIENDKILTDDFSFYVQ	355					
311	DARIERGHSYMGLEFVYFTGDRREYEDVPHMVAVLGPRYRELLDDIERKKKLASDPSIYLH	370					
356	NASASDSDSLAPAGKSALVYLVPMPNNDSGEDQAHQCNVREQVLDLTGLARLGLSDIRAH	415					
371	RPTATDPSMAAGCDTTFVALMPVPHLGSQGDWTWTOAEPYRQSVQEAAL-ERTVLPLGLQHL	429					
416	ECEKIITQWETDEHYVKATFTSLSHKFSQMLYWRPHNRFEELANCYLVGGSTHPGSL	475					
430	RVSFCTTFLDQHRLLSVKAGAGFLEPLLQSAFYRFRPNRSESDVKQLFMWGASTHPGAGV	489					
476	PTIYESAR	483					
490	PGVIMSAK	497					

RESULT 12

352586
phytoene dehydrogenase (EC 1.3.-.-) - *Erwinia herbicola*
?Species: *Erwinia herbicola*
?Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999
?Accession: S52586
?Link, Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.
?J. Gen. Genet. 245, 417-423, 1994

A;title: Transcriptional activation of flanking sequences by Tn1000 insertion.
A;Reference number: S52583; NUID:95107237; PMID:7308390
A;Accession: S52586
A;status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-432 <LIN>
A;Cross-references: EMBL:M90698; NID:G148393; PIDN:AA21263.1; PID:gl48397
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April
C;Superfamily: phytoene dehydrogenase
C;Keywords: oxidoreductase

[illegible]

RESULT 13

A33120
phytoene dehydrogenase (EC 1.3.1.20) - Erwinia herbicola
N:Alternate names: phytoene desaturase
C:Species: Erwinia herbicola
C:Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 29-Sep-1999
C:Accession: A39273; A33120
C:Armstrong, G.A.; Alberti, M.; Hearst, J.E.
Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9978, 1990
A:Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nor
A:Reference number: A39273; MUID:91098634; PMID:2263648
A:Accession: A39273
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-492 <ARM>
A:Cross-references: GB:M38423; NID:G148401; PIDN:AAA24820.1; PID:G148402
A:Note: the authors translated the codon CAG for residue 181 as Phe, TCC for residue 181
A:Note: strain Eho10; ATCC 39368
C:Genetics:
A:Gene: crtI
A:Superfamily: phytoene dehydrogenase

;Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 26.0%; Score 704.5; DB 2; Length 492;

Best Local Similarity 31.7%; Pred. No. 3.7e-45;

Matches 158; Conservative 102; Mismatches 213; Indels 25; Gaps 9;

5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFTDGTFLMKG 64

2 KKTIVGAGFGGLALAIRLQAAGIPTVLEQRDPGRAYVYVHDOGTFDAGPTVITDPT 61

65 VLDENFELCERRSEDYLFELPLSPMYRLLYDDRIIFYV-SDRENWRAELQVDEGTGY 123

62 ALEAFTLAGRMEDYVLLPKPFPYRLCWSGKTLDYANDSABEAQITQFNPRDVEGY 121

124 ECFMEQERKRNALYPCITRDYSSLSK--FLSL-DLIKALP-----WLAPPKSVFNNL 173

122 REFAYSAVQ-----EGYRLGSPVFLSFRDLEAGPQLKLQAW---QSVVQSV 170

174 GQYFNOEKRLAFPCQSKYLGMSWPECALFTMLPYLEHYGIYHVKGGLNRIAAQV 233

171 SRFIDEHLRQAFSPHSLVGNPFTTSIYTLIHALERENGWVFPFGGTGALVNGMVKL 230

234 TAENGEIHLANSIESLIIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPVK 293

231 FTDJGGEIHLARVEELVADNRVSQVRLADGRIFDTAVASNADVNTYKLLGHPVG 290

284 KYTPENLKORYSCSTFMYLGLDKIY-DLPHHTIVFAKYDTNIRNIPONTLTDTSF 352

291 QKRAALERKGSNSLFVLYFLQNPQSOLAHHTICFPYRZELIDRIFTGSAALDDFSL 350

353 YVONASDSDSLAPAGKALVYVPMKN-NDSGLDQWQHCONVREQVLDTLGARLGSDI 411

351 VHLSPCVTDPSLAPPGCASTVLPAPVPLGNAPLDWQBGPKLRDRIPDLYERY-MPGL 409

412 RAHIECEKIITPQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEBELANCYLVGCTHP 471

410 RSQVLTQRIFTPADFDHDLDAHLGSAFSIEPLLTQSAWFRPHNRSDIANLYLVGACTHP 469

472 GSCLPTIYESARISAKLI 489

470 GAGIPGVVASAKATASLM 487

RESULT 14

10061

puaiene synthase (imported) - Staphylococcus aureus (strain N315)

Species: Staphylococcus aureus

Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 02-Aug-2002

Accession: B90061

.. A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

.. A.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

.. not 357, 1225-1240, 2001

Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

Reference number: A89758; UID:21311952; PMID:11418146

Accession: B90061

Status: preliminary

Molecule type: DNA

Residues: 1-502 <KUR>

Cross-references: GB:BA000018; PID:gl3702511; PIDN:BA843652.1; GSPDB:GN00149

Experimental source: strain N315

Genetics:

Gene: crtN

Superfamily: phytoene dehydrogenase

Query Match 25.9%; Score 702.5; DB 2; Length 502;

Best Local Similarity 31.5%; Pred. No. 5.5e-45;

Matches 159; Conservative 108; Mismatches 196; Indels 41; Gaps 12;

7 IIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFTDGTFLMKGYL 66

3 IAVIGAGVTGLAAARASQCHEVTIFEKNNVGRNMQLKDGTFDGMGPTIVMPDVY 62

QY 67 DDMFELCERRSEDYLFELPLSPMYRLLYDDRD-LFVYSDRENWRAELQVDEGTGYEQ 125

DB 63 KDVFACGKNYEDYIEQLQAYIYDVYEDHDDRIYVPTPELAELOQMLSEIIPGSHGFWMS 122

QY 126 FMEQERKRNALYPCITR-----DYSSLSKFSLSLDLIKALPWLAPPKSVFNNL 173

DB 123 FLTDVYKXY-----IARRYFLERTYRKPSDFYNNMTSLVOGAKLKL-----NHA 167

QY 174 GO---YFNOEKRLAFPCQSKYLGMSWPECALFTMLPYLEHYGIYHVKGGLNRIAAA 229

DB 168 DOLIEHYIDNEKIQLLAPOTLYIGIDKRGPSLSYIIPMIEMMEGVHFIKGMVYGAQG 227

QY 230 MAQVTAENGGEIHLANSIESLIIENGAAGVKLOHGAELRG-DEVIINADFAHAMTHLVK 288

DB 228 LAQLNKDLGVNIELNARIEQIITIDPKFRADAIVKNGDIRKFDKILCTADPSPVAESLMP 287

QY 289 P-GVLKCKTTEENLAKQREYSCSTFMYLGLD-KYD-LPHHTIVFAKYDTNIRNIFDNKT 345

DB 288 DFAPIKCKTPEHKAIDLDYSCAFMTIGIDIDVDQVRLHNVISDDPRGNIEIFEGR- 346

QY 346 LTDDFSFYVQNASDSDSLAPAGKALVYVPMK--NNDSGLDW--QAHCONVREQVLD 401

DB 347 LSYDPSIVVYVPAVADKSLAPEKGTIVLPTBELKTGSGIDKSDALTOQIKEIIVRK 406

QY 402 LGARLGSDIRAHIECEKIITPQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEBELAN 461

DB 407 LATIEVFEDIKSHIVSETIFTNDFEQTYHAKPSAFGLMPTLAQSNRYRPNQVSRDYKD 466

QY 462 CYLVGGGTHPGSGLPTIYESARIS 485

DB 467 LYFAGASTPGAGVPIVLTSKIT 490

RESULT 15

D37802

phytoene dehydrogenase (EC 1.3.-.-) crtI - Erwinia uredovora

C:Species: Erwinia uredovora

C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000

C:Accession: D37802

R:Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Harashim

J. Bacteriol. 172, 6704-6712, 1990

A>Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio

A:Reference number: A37802; UID:91072214; PMID:2254247

A:Accession: D37802

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-492 <MIS>

A:Cross-references: GB:D90087; NID:G216681; PIDN:BA414127.1; PID:G216685

C:Superfamily: phytoene dehydrogenase

C:Keywords: oxidoreductase

Query Match 25.8%; Score 700.5; DB 2; Length 492;

Best Local Similarity 31.5%; Pred. No. 7.5e-45;

Matches 157; Conservative 103; Mismatches 213; Indels 25; Gaps 10;

QY 5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFTDGTFLMKG 64

DB 2 KPTTVIGAGFGGLALAIRLQAAGIPTVLEQRDPGRAYVYVHDOGTFDAGPTVITDPS 61

QY 65 VLDENFELCERRSEDYLFELPLSPMYRLLYDDRIIFYV-SDRENWRAELQVDEGTGY 123

DB 62 AIELEFALAGKQLKEYVELLEPTVTFYRLCWESGKVFYNDQDTLEAQIOQFNPRDVEGY 121

QY 124 ECFMEQERKRNALYPCITRDYSSLSK--SFLSL-DLIKALPWL---FPKSVFNNLQ 175

DB 122 RQFLDYSRAVFKGY-----LKLGTVPFLSFRDLEAGPQLKLQAWRSVYSKVAS 172

QY 176 YFNOEKRLAFPCQSKYLGMSWPECALFTMLPYLEHYGIYHVKGGLNRIAAQVIA 235

DB 173 YIEDEHLRQAFSPHSLVGNPFTTSIYTLIHALERENGWVFPFGGTGALVNGMVKL 232

QY 236 ENGGEIHLANSIESLIIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVK--PQWLK 293

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: February 29, 2004, 14:51:24 ; Search time 33.9275 Seconds
(without alignments)
3180.293 Million cell updates/sec

Title: US-09-941-947A-22

Perfect score: 2711

Sequence: 1 MANTKHIIVGAPGGLCAG.....KHVRPKDIAHSANLKAKA 511

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications AA:*
1: /cgn2_6/prodata/2/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/prodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/prodata/2/pubaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/prodata/2/pubaa/US10F_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2711	100.0	511	9	US-09-934-903-16
2	2711	100.0	511	9	US-09-934-868-76
3	2711	100.0	511	10	US-09-941-947A-22
4	2711	100.0	511	14	US-10-358-917-4
5	832.5	30.7	494	15	US-10-369-493-8951
6	816	30.1	517	15	US-10-369-493-19519
7	794.5	29.3	499	15	US-10-369-493-18983
8	767	28.3	498	15	US-10-369-493-20438
9	740	27.3	518	15	US-10-369-493-7750
10	703.5	25.9	492	10	US-09-941-947A-32
11	703.5	25.9	492	14	US-10-218-118-8
12	702.5	25.9	502	14	US-10-358-917-8
13	699.5	25.8	491	15	US-10-369-493-234
14	651.5	24.0	548	15	US-10-369-493-540
15	640.5	23.6	494	9	US-09-547-267-5

Sequence 4, Appli
Sequence 19248, A
Sequence 1769, Ap
Sequence 10579, A
Sequence 19547, A
Sequence 17830, A
Sequence 18, Appl
Sequence 78, Appl
Sequence 24, Appl
Sequence 21541, A
Sequence 8142, Ap
Sequence 18644, A
Sequence 8562, Ap
Sequence 19518, A
Sequence 14, Appl
Sequence 18646, A
Sequence 4194, Ap
Sequence 3, Appli
Sequence 10196, A
Sequence 20440, A
Sequence 5696, Ap
Sequence 5, Appli
Sequence 18, Appl
Sequence 17259, A
Sequence 6, Appli
Sequence 2831, Ap
Sequence 7895, Ap
Sequence 4, Appli
Sequence 23397, A

ALIGNMENTS

RESULT 1

US-09-934-903-16
; Sequence 16, Application US/09934903
; Patent No. US20020102690A1
; GENERAL INFORMATION:
; APPLICANT: Kofas, Mattheos
; APPLICANT: Odum, J. Martin
; APPLICANT: Schenzle, Andreas J.
; APPLICANT: No. US20020102690A1ton, Kelley C.
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Rouviere, Pierre
; APPLICANT: Picataggio, Stephen
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Genes Involved in Isoprenoid Cc pounds Production
; FILE REFERENCE: C11646 US NA
; CURRENT APPLICATION NUMBER: US/09/934,903
; CURRENT FILING DATE: 2001-08-22
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: September 1, 2001
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ORF8

Query Match 100.0%; Score 2711; DB 9; Length 511;
Best Local Similarity 100.0%; Pred. No. 3.8e-270;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANTKHIIVGAPGGLCAGMLISQGFVKVIFDKHAEIG ANRPINMGFTFDGPTPL 60

Db 1 MANTKHIIVGAPGGLCAGMLISQGFVKVIFDKHAEIG ANRPINMGFTFDGPTPL 60

Qy 61 LMKGVLDENFELCERRSELYLEFLPLSEMYRLLYDDRDIF YSDRENMAELQRYFDECT 120

Db 61 LKMGVLDSEFELCERSSEDLPLPLSPMYRLLYDDRDIFVSDRENRAELQRFVDEGT 120
Qy 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180
Db 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180
Qy 181 KRLAFCFQSKYLGMSPWECPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240
Db 181 KRLAFCFQSKYLGMSPWECPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240
Qy 241 IHLNSIESLIIENGAAKGVLQHGAEIRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300
Db 241 IHLNSIESLIIENGAAKGVLQHGAEIRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300
Qy 301 KOREYSCSTFMVLYGMDKXIDYDLPHTTIVPAKDYTNIRNIPENKLTDDFSPYQNASAS 360
Db 301 KOREYSCSTFMVLYGMDKXIDYDLPHTTIVPAKDYTNIRNIPENKLTDDFSPYQNASAS 360
Qy 361 DDLAPAGKSALYVLPVPMNNDGSLDQAHQCNVREQVLTGLARGLSDIRAHIECEKI 420
Db 361 DDLAPAGKSALYVLPVPMNNDGSLDQAHQCNVREQVLTGLARGLSDIRAHIECEKI 420
Qy 421 ITPQTWETDEHYKGTATFSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480
Db 421 ITPQTWETDEHYKGTATFSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480
Qy 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511
Db 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511

RESULT 2

US-09-934-868-76
; Sequence 76, Application US/09934868
; Patent No. US20020137190A1

GENERAL INFORMATION:

; APPLICANT: Koffas, Matteo
; APPLICANT: Odom, James M
; APPLICANT: Schenzle, Andreas J
; TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 76
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Methylobacter 16a
; FEATURE:

OTHER INFORMATION: Amino acid sequences encoded by CRTN1

US-09-934-868-76

Query Match 100.0%; Score 2711; DB 9; Length 511;
Best Local Similarity 100.0%; Pred. No. 3.8e-270;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIQKHAELIRNRPINMNGFTFTGTPTFL 60
Db 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIQKHAELIRNRPINMNGFTFTGTPTFL 60
Qy 61 LKMGVLDSEFELCERSSEDLPLPLSPMYRLLYDDRDIFVSDRENRAELQRFVDEGT 120
Db 61 LKMGVLDSEFELCERSSEDLPLPLSPMYRLLYDDRDIFVSDRENRAELQRFVDEGT 120
Qy 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180
Db 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180
Qy 181 KRLAFCFQSKYLGMSPWECPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240

Db 181 KRLAFCFQSKYLGMSPWECPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240
Qy 241 IHLNSIESLIIENGAAKGVLQHGAEIRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300
Db 241 IHLNSIESLIIENGAAKGVLQHGAEIRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300
Qy 301 KOREYSCSTFMVLYGMDKXIDYDLPHTTIVPAKDYTNIRNIPENKLTDDFSPYQNASAS 360
Db 301 KOREYSCSTFMVLYGMDKXIDYDLPHTTIVPAKDYTNIRNIPENKLTDDFSPYQNASAS 360
Qy 361 DDLAPAGKSALYVLPVPMNNDGSLDQAHQCNVREQVLTGLARGLSDIRAHIECEKI 420
Db 361 DDLAPAGKSALYVLPVPMNNDGSLDQAHQCNVREQVLTGLARGLSDIRAHIECEKI 420
Qy 421 ITPQTWETDEHYKGTATFSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480
Db 421 ITPQTWETDEHYKGTATFSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480
Qy 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511
Db 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511

RESULT 3

US-09-941-947A-22
; Sequence 22, Application US/09941947A
; Publication No. US20030003528A1

GENERAL INFORMATION:

; APPLICANT: Szostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Matteo
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CARBON DIOXIDE PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Methylobacter 16a

US-09-941-947A-22

Query Match 100.0%; Score 2711; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 3.8e-270;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIQKHAELIRNRPINMNGFTFTGTPTFL 60
Db 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIQKHAELIRNRPINMNGFTFTGTPTFL 60
Qy 61 LKMGVLDSEFELCERSSEDLPLPLSPMYRLLYDDRDIFVSDRENRAELQRFVDEGT 120
Db 61 LKMGVLDSEFELCERSSEDLPLPLSPMYRLLYDDRDIFVSDRENRAELQRFVDEGT 120
Qy 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180
Db 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180
Qy 181 KRLAFCFQSKYLGMSPWECPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240
Db 181 KRLAFCFQSKYLGMSPWECPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240

241 IHLNSESLLIENGAAKGVKQHGAEIRGDEVIIINADFAHATHLVKPGVKKYTPENL 300
241 IHLNSESLLIENGAAKGVKQHGAEIRGDEVIIINADFAHATHLVKPGVKKYTPENL 300
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFDMKTLTDDPSFFVQNASAS 360
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFDMKTLTDDPSFFVQNASAS 360
361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVRQVLTGLGRLGSLDIRAHICEKI 420
361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVRQVLTGLGRLGSLDIRAHICEKI 420
421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELA CYLVGGGTHPGSGLPTIYE 480
421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELA CYLVGGGTHPGSGLPTIYE 480
481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511
481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511

RESULT 4
US-10-358-917-4
Sequence 4, Application US/10358917
Publication No. US20030182687A1
GENERAL INFORMATION:
APPLICANT: Cheng, Qiong
APPLICANT: No. US20030182687A1ton, Kelley C.
APPLICANT: Tao, Luan
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS
FILE REFERENCE: CL1929 US NA
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: 60/355,939
PRIOR FILING DATE: 2002-02-11
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 511
TYPE: PRT
ORGANISM: Methylomonas sp.16a
S-10-358-917-4

Query Match 100.0%; Score 2711; DB 14; Length 511;
Best Local Similarity 100.0%; Pred. No. 3,8e-270;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANTKHIIIVGAGPGGLCAGMLLSQRFKVSIFDKHAEIGGRNRPINMNGPTFDGTPTFL 60
1 MANTKHIIIVGAGPGGLCAGMLLSQRFKVSIFDKHAEIGGRNRPINMNGPTFDGTPTFL 60
61 LMKGVLDENFELCERSESDYLEFLPLSPMYRLLYDDRDIFVYSRENKRAELQRFDEGT 120
61 LMKGVLDENFELCERSESDYLEFLPLSPMYRLLYDDRDIFVYSRENKRAELQRFDEGT 120
121 DGYEQMEQERKFNALYPCITRDYSSLSKFLSLDLIKALPLAPPKSVFNNLGOYFNQ 180
121 DGYEQMEQERKFNALYPCITRDYSSLSKFLSLDLIKALPLAPPKSVFNNLGOYFNQ 180
181 KMLAFCFQSKYLGMSWPCPALFTMLPYLSEHYGIYHVKGGLNRIAAQAQVIAENGGE 240
181 KMLAFCFQSKYLGMSWPCPALFTMLPYLSEHYGIYHVKGGLNRIAAQAQVIAENGGE 240
241 IHLNSESLLIENGAAKGVKQHGAEIRGDEVIIINADFAHATHLVKPGVKKYTPENL 300
241 IHLNSESLLIENGAAKGVKQHGAEIRGDEVIIINADFAHATHLVKPGVKKYTPENL 300
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFDMKTLTDDPSFFVQNASAS 360
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFDMKTLTDDPSFFVQNASAS 360
361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVRQVLTGLGRLGSLDIRAHICEKI 420

Db 361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVRQVLTGLGRLGSLDIRAHICEKI 420
Qy 421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELA CYLVGGGTHPGSGLPTIYE 480
Db 421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELA CYLVGGGTHPGSGLPTIYE 480
Qy 481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511
Db 481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511

RESULT 5
US-10-369-493-8951
Sequence 8951, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEIN IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 8951
LENGTH: 494
TYPE: PRT
ORGANISM: Chloroflexus aurantiacus
US-10-369-493-8951

Query Match 30.7%; Score 832.5; DB 15; Length 494;
Best Local Similarity 36.9%; Pred. No. 1.6e-76;
Matches 185; Conservative 97; Mismatches 189; Indels 31; Gaps 12;

Qy 7 IIVGAGPGGLCAGMLLSQRFKVSIFDKHAEIGGRNRPINMNGPTFDGTPTFLMKGV 66
Db 5 IIVGAGPGGMATAIRLAGQGYQVEIFRAVDFGGRMGF LGDYHFDGTPTLQVPRVY 64
Qy 67 DEMFELCERSESDYLEFLPLSPMYRL-LYDDRDIFVYSDE NMRALQRFDEGTG-YE 124
Db 65 DELFSSAGLRFSDYVTLIRLDPNTRIRFDWGEYLDLTSNI AFKAQLAR-FDPALPAEPE 123
Qy 125 Q-FMEQERKFNALYPCITRDYSSLSKFLSLDLIKALPLAPPKSVFNNLGOYFNQ 179
Db 124 RWIEHIRKNVGYEYLGTPVRSPLVYLKPREIAALAF PW----ESLDHFKRFRD 179
Qy 180 EKMRLAFQSKYLGMSWPCPALFTMLPYLSEHYGIYHV GGLNRIAAQAQVIAENGGE 239
Db 180 ERYVYALSYQAKYLGMPHPTACSSVSLVTFLEFADGIVHP GPRALAAGLAKAATDLGV 239
Qy 240 EHLNSESLLIENGAAKGVKQHGAEIRGDEVIIINADFAHATHLVKPGVKKYTPEN 299
Db 240 VHYNSFVQVLEGRACGVELGGERIKADAVVYNADF EALHTIIPPHARGYTQOK 299
Qy 300 LKOREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTNIR-----IPDNKTLTDD 349
Db 300 LNSMEFSCSTFMYLGVNRRWDELPHHQLYL----SANIR NDPLWAKSAVILDE---DD 351
Qy 350 FGFYVONASASDDSLAPAGKSALYVLVMPNNDGLDQAH CONVRQVLTGLGRLGSL 409
Db 352 PSFYVCNPTVDPTNAPGHSFLVLPVPLNRLVPVDDWAA EQRYEDLIIRQM-AKLGFE 410
Qy 410 DIRAHICEKIITPQWETDEHYKATFSLSHKFSQMLYWRPHNRFBELA CYLVGGGTHPGSGLPTIYE 469
Db 411 DVERHVVVERRYTAETWDEHYKATFSLSHKFSQMLYWRPHNRFBELA CYLVGGGTHPGSGLPTIYE 470
Qy 470 HPGSGLPTIYESARISAKLISQ 491

Db 471 HPGSGLMTLEAAKSAVHFIQE 492

RESULT 6

US-10-369-493-19519

Sequence 19519, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 19519

LENGTH: 517

TYPE: PRT

ORGANISM: Myxococcus xanthus

US-10-369-493-19519

Query Match 30.1%; Score 816; DB 15; Length 517;

Best Local Similarity 35.7%; Pred. No. 8.4e-75;

Matches 176; Conservative 106; Mismatches 203; Indels 8; Gaps 6;

2Y 5 KHIIVGAGPGGLCAGMLLSQGRFKVSIIFDKHABIGGRPINNNGFTFTDGTFTFLMKG 64

Db 8 RRIIVVGAGVGLAAARLAHQDFQVFXETQGPGRNCRLQVGDGTDLGPTIVLMPZ 67

2Y 65 VLDEMFCERSESDYLEFLPSPYRLLY-DDRDIFVSDRENRAELQRFDEGTDGY 123

Db 68 VFEETFRVAGRIEDYLLRCDPNYRVHFRDGSVDVFTSELCAWGRELSEVPVGSYARY 127

2Y 124 EQFMEQERKRF-NALYPCITRDYSSLSKSFSLDLIKALPWLAPPKSVFNMLQYFNOEXM 182

Db 128 LAFLAQGRVQYRTSLDLHVGNYAGRLDYLSPVLARIQVRAHRMYADVSPFQDERL 187

2Y 183 RLAFQFSQKYLGMSPWECALFTMLPYLEHYGIVHYVKGGLNRIIAAMAQVIAENGSEIH 242

Db 188 RAATFTQTYLVGVSPYASPAVYGLLPTELGVGIWFPKGGLYATPOALERELAREEGVRFH 247

2Y 243 LNSHIESLIEINGAAKGVKQGAELRGDEVINADFAHAMTHLVKPGVLKXYTPENLXQ 302

Db 248 YGAPVERILTDGRTGVRLEGGVEADAVLGNADLPYAEKLLDP---KATILKREK 304

2Y 303 REYSCSTFMLYGLDKIY-DLPHTTIVPAKYDTNIRNIFDNKTLTDDFSFYVONASASD 361

Db 305 LRYTSSGYMLYLGMRKRYPELLHNHNVFGRDYKSGPDDIFERFRVPEDFSFYVNAPTRTD 364

2Y 362 DSLAPAGKALYLVPMNDSGLDQWACHQVREQVLDTLGRLGLSDIRAHICEKLI 421

Db 365 ASLAPGKDALYLVVPHQHPDLDMKVEGKVRKAFKARN-RELGPFSLESIDIEVERVF 423

2Y 422 TPQWTEDHYVKGATFSLSHKFSQMLYRPHNRFEELANCYLVGGTTHPGSGLPTIYES 481

Db 424 TPDDWAGTFNLARGSAFGLSQNFQIGFFRPSNQDARVKNLFPVGASTDGTGLPTVLIS 483

2Y 482 AR-ISAKLISOKH 493

Db 484 ARLVTERLATWAAH 496

RESULT 7

US-10-369-493-18983

Sequence 18983, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 18983

LENGTH: 499

TYPE: PRT

ORGANISM: Anabaena PCC7120

US-10-369-493-18983

Query Match 29.3%; Score 794.5; DB 15; Length 499;

Best Local Similarity 35.4%; Pred. No. 1.3e-72;

Matches 177; Conservative 102; Mismatches 200; Indels 21; Gaps 10;

QY 4 TKHIIVGAGPGGLCAGMLLSQGRFKVSIIFDKHABIGGRPINNNGFTFTDGTFTFLMK 63

Db 2 SKKVAIVGAGPGGLATAIRLAGLVQVEIFEAERVGGRGFEVDSYAPDTGPTILQLP 61

QY 64 GVLDEMFCERSESDYLEFLPSPMYRL-LYDDRDIFVSDRENRAELQRFDEGTDG 122

Db 62 HLYKELFEEAGFNADYVQLKRLFEYTLKFWDTGTDITLSDLSFKTLQATLERSDPLA 121

QY 123 YEQFMEQERKRFNALY-PCITRDYSSLSKSFSLD-LIKA---PWLAPPKSVFNMLQYF 177

Db 122 FDRWYSEHRIKVELGYKPYLAGPARSIFGYLRPDLAMKFSEFRW---ENLYQHPWREFF 177

QY 178 NOEWRRLAFQSKYLGMSPWECALFTMLPYLEHYGIVHYVKGGLNRIIAAMAQVIAEN 237

Db 178 ODERLVYALSPSKYLGMPHTVASSVSFLIPFLFESQGVHPVGGFEALAGLANAAQDL 237

QY 238 GGEIHNSIEBSLIEINGAAKGVKQGAELRGDEVINADFAHAMTHLVKPGVLKXYTP 297

Db 238 GVKHLHSPVQVWIEQGVKGLADASHRQFTVTVINDFAYAVRHLLFTSARGRYTD 297

QY 298 ENLKOREYSCSTFMLYGLDKIY-DLPHTTIVPAKYDTNIRNIFDNKTLTD-DFS 351

Db 298 NKLQWQFSCSTFMLYGLNRRYEDLPHQIYL----SDLRLEPWWDDSDALDETDP 353

QY 352 FYVONASASDLSLAPAGKALYLVPMNDSGLDQWACHQVREQVLDTLGRLGLSDI 411

Db 354 FYVCPNPTIIDSAPAGHSTFLVLPINTSYAVDMDIKKSTYDFILKRLHL-LGHNI 412

QY 412 RAHICEKLIPTQWTEDHYVKGATFSLSHKFSQMLYRPHNRFEELANCYLVGGTTHP 471

Db 413 BOHIVTQSCYTAQSWLDDYRVHLGAVENLSHNLTLQGPTEPIRSENIAGLYVIGGAVHP 472

QY 472 GSGLPPTIYESARISAKLISQ 491

Db 473 GSGLLTIEASRSAAAGFIHQ 492

RESULT 8

US-10-369-493-20438

Sequence 20438, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

```

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20438
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Rhodopseudomonas palustris
; US-10-369-493-20438

Query Match      28.3%; Score 767; DB 15; Length 498;
Best Local Similarity 32.1%; Pred. No. 8.9e-70;
Matches 157; Conservative 105; Mismatches 221; Indels 6; Gaps 5;

; 6 HIIIVAGPGGLCAGMLLSORGFKVSIFDKHAEIGRNEPIN:NGFTDTGPTTLLMKGLVD 65
; 12 HAVVIGSGFGGLAAAVELGAKGVYVLEKLDKAGGRAYVHQDGFSPAGTIVTAPYL 71
; 66 IDENFELCERRSEDYLEFLPLSPMYRLLYDDRDIF-VYSDRE MDAELQRFDEGTGDEYEQF 124
; 72 FEELMKLIGRMSDDITLKPMSPFYRIFRDDGTHFDYSDRDRAVLQIAKFCDDVPAYD 131
; 125 QFWEQERKRFNALPCITRDYSSLSKLSFLSL-DLIKALP---W AFPKSVFNNLGOYFNQEKM 182
; 132 RFMAASHEIEKVGFEQLGDQPPSHFTDMLK--IAPAMIKLESYRSYGLVAKHFDPKLR 189
; 184 LACFOGSKYLGMSWPCPALFTMLPYLEHEYGIYHVKGGMRIIAAAMAQVIAENGGEIH 243
; 190 QVESFPHLLGGNPFMASSVYCLITYLEKQWGHVSAMGGTALVTGLVNLIEGQGTIRY 249
; 244 NSIESLIIENGAAKGVKLOHGAELRGDEVIINADFAHATHLVKPGVLKXYTPENLKO 303
; 250 NQDVRQIVWENGACGVKLDGSGVIKADIVWSNADASTYRYLLPPETKRWTDAKIERK 309
; 304 EYSCSTFMYLGLDKTY-DLPHTTIVPAKDYTTNIE IFDNKTLTDDFSFYQNAS 358
; 310 RYMSLFWVYFGTKRAYEDVGHHTILLGPRYKELISDFSRKVVAEDFSLYLRHTATDP 369
; 363 SLAPAGKSAIYLVPMNPN--NDSGLDWAHQCNVRE VLDTL-----GALGLS 409
; 370 SLAOGCCTFYLVSPVNLGDTDMTKAETRYASTAKMLGATV-LPDLNQIATSKITT 428
; 423 POTWETDERVYKGTATSLSHKFSQMLYMRPHNRFELANCYLWGGTHPGSLPTIYBSA 482
; 429 PIDFQDLSSFRGAAGLEPVLWQSAWFRPNQSESDVKELYLIVGAGTHPGAGLPGLVLSA 488
; 483 RISAKLISQ 491
; 489 RVLDALVPE 497

RESULT 9
S-10-369-493-7750
Sequence 7750, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
LENGTH: 518
TYPE: PRT

```

```

; ORGANISM: Rhodobacter sphaeroides
; US-10-369-493-7750

Query Match      27.3%; Score 740; DB 15; Length 518;
Best Local Similarity 33.7%; Pred. No. 5.8e-67;
Matches 171; Conservative 96; Mismatches 207; Indels 34; Gaps 10;

; 8 IIVGAGPGGLCAGMLLSORGFKVSIFDKHAEIGRNEPIN:NGFTDTGPTTLLMKGLVD 67
; 14 LVIGSGGLGGLAAAVELGAKGVYVLEKLDKAGGRAYVHQDGFSPAGTIVTAPYL 73
; 68 EMFELCERRSEDYLEFLPLSPMYRLLYDDRDIF-VYSDRE MDAELQRFDEGTGDEYEQF 126
; 74 DLMTKCRDGDADVELKPIDPFYEVWRPDCGSHFTVRQSTE MKAELVARLSPGDVAGYEKF 133
; 127 MEQERKERNALPCITRDYSSLSKLSFLSL-DLIKALP---W AFPKSVFNNLGOYFNQEKM 182
; 134 LKDERKRWYGYEDLGR-----RSMHGKMDLIKVLPTFGM RADRSVYQHAALRVKXDERL 188
; 183 RLAPCFOSKYLGMSPWPCPALFTMLPYLEHEYGIYHVKGGMRIIAAAMAQVIAENGGEIH 242
; 189 RMALSFPHLPIGGDPFNVTSMYILVSQLKEKEFGVHYAIGC AIAAAMAQVIEGQGSFR 248
; 243 LNSEIESLIIENGAAKGVKLOHGAELRGDEVIINADFAHATHLVKPGVLKXYTPENLKO 302
; 249 MNTVEDSILVERKGTATGVRLASGEVLRAGLVSNADAGHT YELLRNHPRRRTDARVKS 308
; 303 RYSCSTFMYLGLDKTY-DLPHTTIVPAKDYTTNIE IFDNKTLTDDFSFYQNAS 358
; 309 RYMSGLFWVYFGTKGKGMWPDVGHTIVNAPRYKGLVE IFLKGLAKMDLSYIHRPS 368
; 359 ASDSLAPAGKSAIYLVPMNPN--NDSGLDWAHQCNVRE VLDTL-----GALGLS 409
; 369 ITDPTVAPEGDDTFYALSPVPHLKQAPVDWQVAEPYRE VLEVLEQSMRPGIGERIGFS 428
; 410 DIRAHECEKIIITPQWETDEHYK-GATFSLSHKFSQML YRPHNRFELANCYLWGG 468
; 429 -----LVFTPEFR-DRLSPNGAGFSTEPRILOSAIFRPHNISEVANLFLVGAG 478
; 469 THPGSLPTIYESARISAKLISQKHRVR 496
; 479 THPGAGVPGVIGSAEVMKAPADAPRAR 506

RESULT 10
US-09-941-947A-32
Sequence 32, Application US/09941947A
Publication No. US20030003528A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia C.
APPLICANT: Cheng, Qiong
APPLICANT: DiCosimo, Deana J.
APPLICANT: Koffas, Mattheos
APPLICANT: Miller, Edward S. Jr.
APPLICANT: Odom, J. Martin
APPLICANT: Picataggio, Steve
APPLICANT: Rouviere, Pierre E.
TITLE OF INVENTION: CARCINOID PRODUCTION FROM A SINGLE CARBON SOURCE
FILE REFERENCE: CL1903 US NA
CURRENT APPLICATION NUMBER: US/09/941,947A
CURRENT FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: 60/229,907
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,858
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Pantoea stewartii
; US-09-941-947A-32

```

Query Match 25.9%; Score 703.5; DB 10; Length 492;
 Best Local Similarity 31.7%; Pred. No. 3.1e-63;
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFDGPTFLMKG 64
 DB 2 KPTTVIGAGGGLALAIRLQAAGIPVLLLEQDRKPGGRAYVYQGGFTFDAGPTVITDPS 61
 QY 65 VLDEMFCERSSDYLFPLSPMYRLLYDDRDIFVY-SDENNRAELQRFVDEGTDGY 123
 DB 62 ALIEELFALAGLQKDYVELLPVTFYRLCWESGKVFYNDQAOLEAQIOQFNPRDVAGY 121
 QY 124 EQFMEQERKRNFALPCITRDYSSLK-----SFLSL-DLIALPWLIA---PFSKVFNNLQ 175
 DB 122 RAFLDYSRAVNEG-----LKLGTVPFLSKDMLRAAPQAKLQAWRSVYSKVG 172
 QY 176 YFNQEMRLAFCQSKYLGMSWPCPALFTMLPYLEHEYGIIHVKGGLNRIAAQAQVIA 235
 DB 173 YIEDEHLQAQSFSLVGGNPFATSSIIYTLIHALEREWVMPRGCTGALVNGMKLFQ 232
 QY 236 ENGGEIHLNSIESLIIENGAAGKVKLGQGAELRGDEVIINADPAHAMTHLVK--PGVLK 293
 DB 233 DLGGEVNLARVSHMETVGDKIQAQVLEDGRFRFTCAVSNADVVHYTRDLLSQHPAAK 292
 QY 294 KYTPENLQREYSCTFMYLGLDKIYD-LPHHTIVFAKDYTNIRNIPFNKTLDTPSP 352
 DB 293 Q--AKKLQSKMSNSLFVYFGLNHHDDQLAHTVCFGRYRELIIHFHHDGLAEDFSL 350
 QY 353 YVQNASASDDSLAPAGKSAVYVMPMN-NDSGLDQAHQNVREQVLDTLGARGLSDI 411
 DB 351 YLHAPCVTPSLAPBGCGSYVYVAPVPHLGTANLDMAVEGRLDRIYDFYLEQHY-MPGL 409
 QY 412 RAHIECEKLIIPQWETDEHYKGAATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHP 471
 DB 410 RSQVTHRMFTFPDRDELNAWQSAFSEVPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
 QY 472 QSGLPITYESARISAKLI 489
 DB 470 GAGIPGVIGSAKATAGLM 487

RESULT 11
 US-10-218-118-8
 ; Sequence 8, Application US/10218118
 ; Publication No. US20030148319A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bzostowicz, Patricia
 ; APPLICANT: Rouviere, Pierre
 ; APPLICANT: Picataggio, Stephen
 ; APPLICANT: Cheng, Qiong
 ; TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
 ; FILE REFERENCE: CL1876 US NA
 ; CURRENT APPLICATION NUMBER: US/10/218,118
 ; CURRENT FILING DATE: 2002-08-13
 ; PRIOR APPLICATION NUMBER: 60/312,646
 ; PRIOR FILING DATE: 2001-08-15
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 8
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Pantoea stewartii
 US-10-218-118-8

Query Match 25.9%; Score 703.5; DB 14; Length 492;
 Best Local Similarity 31.7%; Pred. No. 3.1e-63;
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFDGPTFLMKG 64
 DB 2 KPTTVIGAGGGLALAIRLQAAGIPVLLLEQDRKPGGRAYVYQGGFTFDAGPTVITDPS 61
 QY 65 VLDEMFCERSSDYLFPLSPMYRLLYDDRDIFVY-SDENNRAELQRFVDEGTDGY 123

DB 62 ALIEELFALAGLQKDYVELLPVTFYRLCWESGKVFYNDQAOLEAQIOQFNPRDVAGY 121
 QY 124 EQFMEQERKRNFALPCITRDYSSLK-----SFLSL-DLIALPWLIA---PFSKVFNNLQ 175
 DB 122 RAFLDYSRAVNEG-----LKLGTVPFLSKDMLRAAPQAKLQAWRSVYSKVG 172
 QY 176 YFNQEMRLAFCQSKYLGMSWPCPALFTMLPYLEHEYGIIHVKGGLNRIAAQAQVIA 235
 DB 173 YIEDEHLQAQSFSLVGGNPFATSSIIYTLIHALEREWVMPRGCTGALVNGMKLFQ 232
 QY 236 ENGGEIHLNSIESLIIENGAAGKVKLGQGAELRGDEVIINADPAHAMTHLVK--PGVLK 293
 DB 233 DLGGEVNLARVSHMETVGDKIQAQVLEDGRFRFTCAVSNADVVHYTRDLLSQHPAAK 292
 QY 294 KYTPENLQREYSCTFMYLGLDKIYD-LPHHTIVFAKDYTNIRNIPFNKTLDTPSP 352
 DB 293 Q--AKKLQSKMSNSLFVYFGLNHHDDQLAHTVCFGRYRELIIHFHHDGLAEDFSL 350
 QY 353 YVQNASASDDSLAPAGKSAVYVMPMN-NDSGLDQAHQNVREQVLDTLGARGLSDI 411
 DB 351 YLHAPCVTPSLAPBGCGSYVYVAPVPHLGTANLDMAVEGRLDRIYDFYLEQHY-MPGL 409
 QY 412 RAHIECEKLIIPQWETDEHYKGAATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHP 471
 DB 410 RSQVTHRMFTFPDRDELNAWQSAFSEVPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
 QY 472 QSGLPITYESARISAKLI 489
 DB 470 GAGIPGVIGSAKATAGLM 487

RESULT 12
 US-10-358-917-12
 ; Sequence 12, Application US/10358917
 ; Publication No. US20030182687A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cheng, Qiong
 ; APPLICANT: No. US20030182687A1ton, Kelley C.
 ; APPLICANT: Tao, Luan
 ; TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOI COMPOUNDS
 ; FILE REFERENCE: CL1929 US NA
 ; CURRENT APPLICATION NUMBER: US/10/358,917
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: 60/355,939
 ; PRIOR FILING DATE: 2002-02-11
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 12
 ; LENGTH: 502
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-10-358-917-12

Query Match 25.9%; Score 702.5; DB 14; Length 502;
 Best Local Similarity 31.5%; Pred. No. 4e-63;
 Matches 159; Conservative 108; Mismatches 196; Indels 41; Gaps 12;

QY 7 IIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFDGPTFLMKGV 66
 DB 3 IAVIGAGVTGLAAARIASQGEHVTFEKNNVGRMNQKXQGTFTDMGPTIVMPEVY 62
 QY 67 DEMFELCERSSDYLFPLSPMYRLLYDDRD-IFVYSD-ENMRAELQRFVDEGTDGYE 125
 DB 63 KDVFTACGKNVEDYIELQRLAYIVDYFDHDDRTVPTD-AELQQLMESIEPSTHGFMS 122
 QY 126 FMEQERKRNFALPCITR-----DYSLKSL-LDLIKALPMLAPFKSVFNNL 173
 DB 123 FLTDVYKYE-----IARRYFLERTYRKPSDFYNNYSLVJAKUKTL-----NHA 167
 QY 174 GQ-----YFNQEMRLAFCQSKYLGMSWPCPALFTMLP-LEHEYGYHVKGGLNRIAAA 229
 DB 168 DQLEIHYIDNEKIQLAFQTLIYIGIDPKRGPSLYSIIP-LEMMFGVHFHFKGMYGNAQG 227

429 DEHYKATGATSLSHKTSOMLYWRPHNRFBELANCYLVGCTHPTGSGLPITYESARISAKL 488
476 TLISYLGNAFGPEPRVLQSAFFPHNRSEDLHNFLVYGAGAQPGAGTSPVMSAKTARL 535
489 ISQ 491
536 IAE 538

RESULT 15
US-09-547-267-5
Sequence 5, Application US/09547267
Patent No. US20020147371A1
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

Query Match 23.6%; Score 640.5; DB 9; Length 494;
Best Local Similarity 30.9%; Pred. No. 9.7e-57;
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;
8 IIVGAGPGCLCAGMLLSQGFVKSIIDKHAIEIGGRNPINWNGFTDTGPTLLMKGVLD 67
5 IIVGAGFGLALAIKQAGATTIVARDKPGGRAYVWDQGHVFDAGFTVVDPSLR 64
68 EMFELCERRSELYLEFLSPMYRLYDDRDIIFYV-SDRENMRALQRFDEGTDGYEQF 126
65 ELWALSGQPMERDVTLLFVSPFYRLTWADGRSPFYVNDDELIRQVASPNPADVDGYRRF 124
127 MEQERKRNFALYPCIT-TRDYSLSKSLFSLDLIKALPWLAPPKSVFNMLGOYFNQEXMLA 185
125 HDYAEVYREGYKLGGITTFP--LKLQOMLNAAPALMELQAYRSVHSMVAFIOPDHLRQA 182
186 FCFQSKYLGMSPECPALFTMLPYLEHEYGIYHKVGLNRIIAAQAQVIAENGGEIHLNS 245

Db 183 FSHFTLLVGNPFSTSSIVALTIALERROGWWFAKGGTNLVAGMVALFERLGGTILLNA 242
QY 246 EIESLIIEENGAAGVKLQHCAGELRGDEVIINADPAHAMTI LKPGVLKKYTPENLKOREY 305
Db 243 RVTIDTEGDRATGVTLIDGRQLRADTVASNGDVMSYRI LLGHTRRGRTKAAILNRQRM 302
QY 306 SCSTFMLYLGLDK-IYDLPHHTIVPAKDYTTNIRNIFDNVLTDDPSFYVONASASDDSL 364
Db 303 SMSLFVLHFLGSKRPENLAHSHVIFGPRYKGLVNEIFNGERLPDDFSMYLHSPCTVDPSL 362
QY 365 APAGKSALYVLVPMNP-NDSGLDWQAHQCNVREQVLTLCARLGLSDIRAHIECEKIITP 423
Db 363 APEGMSTHYVLAVPVHLGRADVWEAEAPGYAERIPEEL-ERRAIPDLAKHLTVSRIFSP 421
QY 424 QTWETDEHYKATGATFSLSHKFSOMLYWRPHNRFBELANC LKCGGTHPTGSGLPITYESAR 483
Db 422 ADFSTELSAEHGSAFSEVPEPILQTSANFRPHNRDRALPNF IVGAGTHPGAGIPGVVGSAR 481
QY 484 ISAKLI 489
Db 482 ATAQVM 487

Search completed: February 29, 2004, 15:27:50
Job time : 37.9275 secs

GenCore version 5.1.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 15.5283 Seconds
(without alignment)
1698.885 Million cell updates/sec

file: US-09-941-947A-22

effect score: 2711

sequence: 1 MANTKHIIIVGAGPGGLCAG.....KRVSRFKDIAHSAWLKAKA 511

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA.*

1: /cgn2_6/prodata/2/iaa/5A COMB.pap.*

2: /cgn2_6/prodata/2/iaa/5B COMB.pap.*

3: /cgn2_6/prodata/2/iaa/6A COMB.pap.*

4: /cgn2_6/prodata/2/iaa/6B COMB.pap.*

5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pap.*

6: /cgn2_6/prodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2711	100.0	511	4	US-09-934-903-16
2	700.5	25.8	492	1	Sequence 16, Appli
3	673	24.8	582	1	Sequence 4, Appli
4	640.5	23.6	494	3	Sequence 17, Appli
5	640.5	23.6	494	3	Sequence 5, Appli
6	640.5	23.6	494	3	Sequence 5, Appli
7	640.5	23.6	494	3	Sequence 5, Appli
8	640.5	23.6	494	3	Sequence 5, Appli
9	640.5	23.6	494	3	Sequence 5, Appli
10	614	22.6	489	1	Sequence 4, Appli
11	614	22.6	489	1	Sequence 8, Appli
12	614	22.6	489	1	Sequence 8, Appli
13	611	22.5	489	1	Sequence 10, Appli
14	611	22.5	489	1	Sequence 10, Appli
15	611	22.5	489	1	Sequence 10, Appli
16	607.5	22.4	497	4	Sequence 18, Appli
17	237.5	8.8	290	4	Sequence 462, App
18	206	7.6	610	4	Sequence 55, Appli
19	140.5	5.2	524	3	Sequence 1, Appli
20	140.5	5.2	524	4	Sequence 1, Appli
21	126.5	4.7	588	4	Sequence 13, Appli
22	125.5	4.6	489	4	Sequence 2, Appli
23	121	4.5	354	4	Sequence 958, App
24	119.5	4.4	475	4	Sequence 13710 A
25	119.5	4.4	582	1	Sequence 4, Appli
26	119	4.4	582	1	Sequence 6, Appli
27	118.5	4.4	568	4	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-934-903-16

; Sequence 16, Application US/09934903

; Patent No. 6660507

; GENERAL INFORMATION:

; APPLICANT: Koffas, Mattheos

; APPLICANT: Odum, J. Martin

; APPLICANT: Schenzle, Andreas J.

; APPLICANT: No. 6660507ton, Kelley C.

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Rouviere, Pierre

; APPLICANT: Picataggio, Stephen

; APPLICANT: Cheng, Giong

; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

; FILE REFERENCE: C11646 US NA

; CURRENT APPLICATION NUMBER: US/09/934,903

; CURRENT FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/229,907

; PRIOR FILING DATE: September 1, 2001

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 16

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Methylobionas 16a

; FEATURE:

; OTHER INFORMATION: Amino acid sequences encoded by ORF8

US-09-934-903-16

Query Match 100.0%; Score 2711; DB 4; Length 511;

Best Local Similarity 100.0%; Pred. No. 9,1e-266;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANTKHIIIVGAGPGGLCAGMLLSORGFVSVDFDKHAGIC:RNRPINMGFTDGTPTFL 60

Db 1 MANTKHIIIVGAGPGGLCAGMLLSORGFVSVDFDKHAGIC:RNRPINMGFTDGTPTFL 60

Qy 61 LMGVLDMEFELCERSEDYLEPLSPMYRLLYDDRDIF YSDRENMAELQVDEGT 120

Db 61 LMGVLDMEFELCERSEDYLEPLSPMYRLLYDDRDIF YSDRENMAELQVDEGT 120

Qy 121 DGVEQFMEQERKFNALYPCITRDYSSLSKLSFLSLLDKALPMLAPKPSVFNNGQVFNQE 180

Db 121 DGVEQFMEQERKFNALYPCITRDYSSLSKLSFLSLLDKALPMLAPKPSVFNNGQVFNQE 180

Qy 181 KRLAFCFQSKYLGMSFWPCPALFTMLPYLEHYGYHYVK ELNRIIAAAMAQVIAENGCE 240

Db 181 KRLAFCFQSKYLGMSFWPCPALFTMLPYLEHYGYHYVK ELNRIIAAAMAQVIAENGCE 240

Qy 241 IHLNSEIESLIIENGAAGVKVQHGALRGDEVIINADFA:AMTHLVKPGVLKXKYPENL 300

Db 241 IHLNSELISLIENGAAKGVKLGAEIRGDEVIINADFAAMTHLVKPGVLKXITPENL 300
QY 301 KOREYSCSTFMYLGLDKIYDLPHETIIVFAKYTTNIRNIFDNKTLTDDFSFYVQNASAS 360
Db 301 KOREYSCSTFMYLGLDKIYDLPHETIIVFAKYTTNIRNIFDNKTLTDDFSFYVQNASAS 360
QY 361 DLSLAPAGKSALYLVVPPNDSGLDWAHCONVREOVLDTLGLARGLSDIRAHIECEKI 420
Db 361 DLSLAPAGKSALYLVVPPNDSGLDWAHCONVREOVLDTLGLARGLSDIRAHIECEKI 420
QY 421 ITPQWETDEHYKATGATFSLSHKFSQMLYWRPHNRFEEELANCYLVGSGTHPGSLPTIYE 480
Db 421 ITPQWETDEHYKATGATFSLSHKFSQMLYWRPHNRFEEELANCYLVGSGTHPGSLPTIYE 480
QY 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511
Db 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511

RESULT 2
US-07-783-705A-4
; Sequence 4, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939Ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, ShigeYuki
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ladass & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/783.705A
; FILING DATE: 1991:023
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-103078
; FILING DATE: 21-APR-1989
; APPLICATION NUMBER: JP 2-53225
; FILING DATE: 05-MAR-1990
; APPLICATION NUMBER: US 07/519,011
; FILING DATE: 19-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwadron, Janet I.
; REGISTRATION NUMBER: 33,778
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-708-1935
; TELEFAX: 212-246-5959
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-705A-4

Query Match 25.8%; Score 700.5; DB 1; Length 492;
Best Local Similarity 31.5%; Pred. No. 4.4e-62;
Matches 157; Conservative 103; Mismatches 213; Indels 25; Gaps 10;

QY 5 KHLIIVGAPGGLCAGMLLSQSGFKVSIFFDKHAEITGGRN PINNMGFTFTDPTFLMKG 64
Db 2 KPTTVIGAGFGALAIQLQAAGIPVLLLEQRDPGGHA VYEDQGFTFDAGPIVITDPS 61
QY 65 VLDEMFELCERSEDIYFLPLSPMYRLLYDDRIDFVY- DRENMRAELQRFVDEGTDGY 123
Db 62 AEBELPALAGKQLKEYVELLPVTPPYRLCWESGKVFNVDI DOTRLEAQIQCFNPRDVEGY 121
QY 124 EQFMEQERKRFNALYPCITRDYSSLK-----SFLSL-DLI-ALPWL- ---FPKSVFNILGQ 175
Db 122 ROFLDYSRAVFKEGY-----LKLGTVPFLSFRDMLAAPOLAKICAWRSVYSKVAS 172
QY 176 YFNQEMRLAFQFQSKYLGMSFPCPALFTMLPYLEHEVYIYHVKGGLNRJAAAMAQVIA 235
Db 173 YTEDHLLQAQSFHSLLVGNGPPATSSIVTILHALEREW VWFPRGGTGALVQGMKILFQ 232
QY 236 ENGSIHLNSELISLIENGAAKGVKLGAEIRGDEVI NADFAAMTHLVK--PGVLK 293
Db 233 DLGGEVILNARVSHMETTGNKIEAVHLEDGERFUTQAVA NADVHTYRDLLOSHPAAVK 292
QY 294 KYTPENLKOREYSCSTFMYLGLDKIYD-LPHHTIVFAK YTTNIRNIFDNKTLTDDFSF 352
Db 293 Q--SNKLOTKMSNSLVLYYFGLNHHHDQLAHTVCFGP YRELIIDEIFNHDGLAEDPSL 350
QY 353 YVQNASDSDSLAPAGKSALYLVVPMEN-NDSGLDWAHCONVREOVLDTLGLARGLSDI 411
Db 351 YLHAPCVTDSLSLAPGCGSYVVLAPVPHLGTANLDWTVE PKLRDRIPAYLEQHY-MPGL 409
QY 412 RAHIECEKIITPQWETDEHYKATGATFSLSHKFSQMLYK PHNRFEEELANCYLVGSGTHP 471
Db 410 RSLQVTHRMFTFFPFRDQNLNAYHGSASFVEPVLTQSAWF PHNRDKTITNLYLVGAGTHP 469
QY 472 GSGLPTIYESARISAKLI 489
Db 470 GAGIPGVIGSAKATAGLM 487
RESULT 3
US-09-091-725-17
; Sequence 17, Application US/09091725
; Patent No. 6329141
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Improved methods for transforming Phaffia
; TITLE OF INVENTION: and recombinant DNA for use therein
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster llp
; STREET: 2000 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: United States of America
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.2 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091.725
; FILING DATE: 23-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95203620.0
; FILING DATE: 22-DEC-1995
; APPLICATION NUMBER: EP 96200943.7
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: E. Victor Donahue
; REGISTRATION NUMBER: 35,492
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 582 amino acids

Db 303 SMSLFLVHFLGSKRPENLAHSVIFGPRYKGLVNEIFNGPRLPDDFSKYLHSPCVTDPSL 362
QY 365 APAGKSALYVLVPMEN-NDSGLDQWQHCONVREQVLDLTGRLGLSIRAHIECEKIITP 423
Db 363 APEGMSHYVLAPVPHLGRADVDEARAPGYAERIFBEL-ERRAIPDLRKLITVSRIFSP 421
QY 424 QTWETDEHVYKATFSLSHKFSQMLYWRPHNRPFEELANCYLVGSGTHPGSLPTIYESAR 483
Db 422 ADFSTELSAHGSAPFSEPIILTOSAWFRPHNDRRAIFNFYIVGAGTHPGAGIPGVVGSAAK 481
QY 484 ISAKLI 489
Db 482 ATAQVM 487

RESULT 7
JS-08-980-832-4
; Sequence 4, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tsygankov, Yuri
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
; FILE REFERENCE: Improved Fermentative Carotenoid
; CURRENT APPLICATION NUMBER: US/08/980.832B
; CURRENT FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
JS-08-980-832-4

Query Match 23.6%; Score 640.5; DB 3; Length 494;
Best Local Similarity 30.9%; Pred. No. 5.3e-56;
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;

Y 8 IIVGAGPGGLCAGMLLSQRFKYSIFDKHAEIGGRNRPINANGFTDTGPTFLMKGVLD 67
b 5 IVIGAGFGGLALAIRLQSGAIIIVAEARKPGRAYVNDQGHVFDAGFTVVDPSLR 64
Y 68 EMPFELCRRSDYLEFLPLSPMYRLLYDDRDIFVY-SDRENMPAEIQRVDEGTGVEQF 126
b 65 ELWALSQPMERDVTLPLVSPFFYRLTWADGRSPFYVNDDELIRQVASFNPADVDGYRRF 124
Y 127 MEQERKRFNLYPCI-TRDYSSLSKSFSLDLIKALPWLAFPKSVFNNGQYFNQKRLA 185
b 125 HDYAEVYRGYKLGITTPF--LKQQLMAAPALMRQAIRSVHSMVAFIQQDPLHQA 182
Y 186 FCFOSKYLGMSPWECALFTMLPYLHEHYGIYHVKGGLNRIAAQAQVIAENGGEIHLMS 245
b 183 FSPHTLVGNGPSTSIYALIHALLERRGGVWFAKGGTQVLVAGMVALFERLGGTLLNA 242
Y 246 EIESLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLKYYTPENLKQREY 305
b 243 RVTRIDTEGDRATGVTLDRQRLRADTVASNGDWMHSYEDLLGHTRRGRRTKAAILNRQW 302
Y 306 SCSTFMLYLGLDK-IYDLPHHTIVFAKDYTTNIRNIFDNKLTDDPSFYVQNASDDSL 364
b 303 SMSLFLVHFLGSKRPENLAHSVIFGPRYKGLVNEIFNGPRLPDDFSKYLHSPCVTDPSL 362
Y 365 APAGKSALYVLVPMEN-NDSGLDQWQHCONVREQVLDLTGRLGLSIRAHIECEKIITP 423
b 363 APEGMSHYVLAPVPHLGRADVDEARAPGYAERIFBEL-ERRAIPDLRKLITVSRIFSP 421
Y 424 QTWETDEHVYKATFSLSHKFSQMLYWRPHNRPFEELANCYLVGSGTHPGSLPTIYESAR 483
b 422 ADFSTELSAHGSAPFSEPIILTOSAWFRPHNDRRAIFNFYIVGAGTHPGAGIPGVVGSAAK 481
Y 484 ISAKLI 489
b 482 ATAQVM 487

RESULT 8

US-09-547-267-5
; Sequence 5, Application US/09547267
; Patent No. 6613543
; GENERAL INFORMATION:
; APPLICANT: Hohmann, Hans-Peter
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tessier, Michel
; APPLICANT: van Loon, Adolphus
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: NJ
; COUNTRY: USA
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/547,267
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,645
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pokras, Bruce A.
; REGISTRATION NUMBER: 32,748
; REFERENCE/POCKET NUMBER: RAN 6002/170
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-5801
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-547-267-5

Query Match 23.6%; Score 640.5; DB 4; Length 494;
Best Local Similarity 30.9%; Pred. No. 5.3e-56;
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;

QY 8 IIVGAGPGGLCAGMLLSQRFKYSIFDKHAEIGGRNRPINANGFTDTGPTFLMKGVLD 67
Db 5 IVIGAGFGGLALAIRLQSGAIIIVAEARKPGRAYVWN QGHVFDAGFTVVDPSLR 64
QY 68 EMPFELCRRSDYLEFLPLSPMYRLLYDDRDIFVY-SDRENMPAEIQRVDEGTGVEQF 126
Db 65 ELWALSQPMERDVTLPLVSPFFYRLTWADGRSPFYVNDDELIRQVASFNPADVDGYRRF 124
QY 127 MEQERKRFNLYPCI-TRDYSSLSKSFSLDLIKALPWLAFPKSVFNNGQYFNQKRLA 185
Db 125 HDYAEVYRGYKLGITTPF--LKQQLMAAPALMRQAIRSVHSMVAFIQQDPLHQA 182
QY 186 FCFOSKYLGMSPWECALFTMLPYLHEHYGIYHVKGGLNRIAAQAQVIAENGGEIHLMS 245
Db 183 FSPHTLVGNGPSTSIYALIHALLERRGGVWFAKGGTQVLVAGMVALFERLGGTLLNA 242
QY 246 EIESLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLKYYTPENLKQREY 305
Db 243 RVTRIDTEGDRATGVTLDRQRLRADTVASNGDWMHSYEDLLGHTRRGRRTKAAILNRQW 302
QY 306 SCSTFMLYLGLDK-IYDLPHHTIVFAKDYTTNIRNIFDNKLTDDPSFYVQNASDDSL 364

Db 303 SMELFVLHFLGSLKRPENLAHRSVIFGRYKGLVNEIPNGFRIPDDFMSYLSHSPCVTDPSL 362
QY 365 APAGKSAIYLVMPEN-NDUGLDWQACQVVEQVLDLTGLARGLSDIRAHIEKEKIITP 423
Db 363 APEGMTHTYVLAPVPHLGRADVWEAEAPGYAERIFEL-ERRAIPDLRKLTVSRIFSP 421
QY 424 QTWETDEHYKATFSLSKFSQMLYWRPHNFEELANCYLVGCGTHFGSGLTIFYESAR 403
Db 422 ADPSTELSAHGSASFVEPILOTQSAWFRPHNDRRAIPNFYIVGAGTHFGAGIPGVWGSAX 401
QY 484 ISAKLI 489
Db 482 ATAQVM 487

RESULT 9

US-09-920-923B-4
; Sequence 4, Application US/09920923B
; Patent No. 6677134
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tsygankov, Yuri
; TITLE OF INVENTION: Fermentative Carotenoid Production.
; FILE REFERENCE: 15464 US (C38435/125944)
; CURRENT APPLICATION NUMBER: US/09/920,923B
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 08/980,832
; PRIOR FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Flavobacterium sp. R1534
US-09-920-923B-4

Query Match 23.6%; Score 640.5; DB 4; Length 494;
Best Local Similarity 30.9%; Pred. No. 5.3e-56;
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;
QY 8 IVGAGPGGLGAGMLLSQSGFKVSIPTDKHAEIGGRNRPINNGFTTDTGPTFLMKGVLD 67
Db 5 IVGAGFGGLALAIRLQSGAGIATTIVEARDKPGGRAYVYNDQGHVFDAGPTVTTDPSLR 64
QY 68 EMFELCERSEDEFLPLSPMYRLLYDDRDIFVY-SDRENMAELQRFDEGTDGYEQF 126
Db 65 ELWALSQGPWEDVTLPPVSPYRLTWADGRSFYVNDDELIRQVASFNPADVDGTRRF 124
QY 127 MEQERKRFNALYPCI-TRDYSSLSKSLFLDLIKALPWLAFPKSVFNILQGYFNOERKRLA 185
Db 125 HOYAEVVRREGYLKLTGTPF--LXLGQMLNAPALMRLOQAVRSVHSMVAREIQPHLRQA 182
QY 186 FCFQSKYLGMSWECFALFTMLPYLEHYGYHYKGLNRIAAAMAQVIAENGGEIHLNS 245
Db 183 FGFHTLLVGNPFSTSSYALIHALERGGVWFAGKGNQLVAGVVALFELGCGTLLNA 242
QY 246 EIESIIEENGAAKGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYKXYPENLKOREY 305
Db 243 RVTRIDTEGDRATGVTLLDGRCLRADTVASNGDVMSYRDLHGTRGRKTAALNQRW 302
QY 306 SCSTFMVLYGLDK-LYDLPHITVIFAKDYTNIRNIFDNKLTLDSPFYVONASASDLSL 364
Db 303 SMELFVLHFLGSLKRPENLAHRSVIFGRYKGLVNEIPNGFRIPDDFMSYLSHSPCVTDPSL 362
QY 365 APAGKSAIYLVMPEN-NDUGLDWQACQVVEQVLDLTGLARGLSDIRAHIEKEKIITP 423
Db 363 APEGMTHTYVLAPVPHLGRADVWEAEAPGYAERIFEL-ERRAIPDLRKLTVSRIFSP 421
QY 424 QTWETDEHYKATFSLSKFSQMLYWRPHNFEELANCYLVGCGTHFGSGLTIFYESAR 403
Db 422 ADPSTELSAHGSASFVEPILOTQSAWFRPHNDRRAIPNFYIVGAGTHFGAGIPGVWGSAX 401

QY 484 ISAKLI 489
Db 482 ATAQVM 487
RESULT 10
US-08-095-726-8
; Sequence 8, Application US/08095726
; Patent No. 5530188
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L
; APPLICANT: Brinkhaus, Friedhelm L
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H
; APPLICANT: Yarger, James G
; APPLICANT: Yen, Huel-Che B
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in
; TITLE OF INVENTION: Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept
; STREET: 200 E Randolph St
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,726
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/785,566
; FILING DATE: 30-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5530188val B
; TELEPHONE: 3128567180
; TELEFAX: 3128564972
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-095-726-8

Query Match 22.6%; Score 614; DB 1; Length 489;
Best Local Similarity 30.3%; Pred. No. 2.5e-53;
Matches 152; Conservative 97; Mismatches 210; Indels 34; Gaps 12;
QY 5 KHIIVGAGPGGLGAGMLLSQSGFKVSIPTDKHAEIGGRNRPINNGFTTDTGPTFLMKG 64
Db 2 KKTWVIGAGFGGLALAIRLQAGIPTVLLEQDKPGGRAYVHMQGPTFDAGPTVITDPT 61
QY 65 VLDEMFCERSEDEFLPLSPMYRLLYDDRDIFVY-SDRENMAELQRFDEGTDGY 123
Db 62 ALAELFTLAGRMEDYVRLPVKPFYRLCWESGKTLOYA-DSFELEQITQFNPRDVEGY 121
QY 124 EGFMEQERKRFNALYPCI-TRDYSSLSKSLFLSL-DLIKALP-----WLAFPKSVFNIL 173
Db 122 RRLFAYSQAVFQ-----EGYLRGVSVPFLSRDMLRAQQLKLQAW-----QSVYQSV 170
QY 174 GYFQNBKMLAFCFQSKYLGMSWECFALFTMLPYLEHYGYHYKGLNRIAAAMAQV 233
Db 171 SFPIEDEHLRQAFSPHSLVGNPPTSSIVYLHALERMGVAFPEGGTGALVNGWKL 230
QY 234 IAEENGIEHLNSEIESLIENGAAKGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYK 293

467 GAGIPGVVGLAESTASLM 484

RESULT 14

IS-08-096-043-10
Sequence 10, Application US/08096043
Patent No. 5530189

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Lycopene Biosynthesis in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,043
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,568
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

IS-08-096-043-10

Query Match 22.5%; Score 611; DB 1; Length 489;
Best Local Similarity 30.3%; Pred. No. 5e-53;
Matches 151; Conservative 97; Mismatches 216; Indels 34; Gaps 12;
8 IIVGAGPGLCAGMILLSQRFKVSIFDKHAEIGGRNRPINMGFTPTDGTFLAKGVLD 67
5 WVIGAGPGLALAIRLQAAGIPTVLEQREKPGRAYVWHDQGTFTDAGTPTITDPTALE 64
68 EMPELCERSSELYLEPLSPMYRLLYDDRDIFVY--SDRENMAELQRVDSGTGYBQF 126
65 ALFTLAGRMEDVRLLPVAPFFRLCWESKTLVDYANDSFELAQITQNPDDVGGIRFP 124
127 MEQERKFNALYPCITRDYSLLKS--PLSL-DLIKALP-----WLAPPKSVFNNGOY 176
125 LAYSQAVFQ-----EGYRLGVSVPFLSFRDMLRAGPQLKLQAW---QSVYQSVSRF 173
177 FNQEKMLAFQSKYLGMSFWSCPALFTMLPLYEHEYGYHYVKGGLNRIIAAQAQVIAE 236
174 IEDHLEQAFSFTSLVGGNPFPTSSYITLIHALEREWGVWPPGGTGAIVNGVYKLFDT 233
237 NGGEIHLNSESIESLIENGAAKGVKQHGAEELRGDEVIINADPAHAMTHLVKPGVLYKYT 296
234 LGGEIELNARVELVWADNRVSRQLADGRIFPTDVAASADWNTYKKLGLGTIPVGQKE 293

QY 297 PENLKQREYSCSTFMYLGLGLKLY-DLPHTTIVFAKDYTTNIRNIFDNKTLTDDPSFYVQ 355
DB 294 AARLERXSMNSLFLVYFGLNQPHSQLAHTTICGPRYRELIDEITGSGALADDFSLYLH 353
QY 356 NASASDDSLAPAGKALYLVPMNP-NDGLDQQAHCQNVREQVLDTLGARLGLSIRAH 414
DB 354 SPCVTDPSLAPPCASFYVLAPVPHLGNAPLDWAQEGPKLRDRIFDYLEERY-MPGILRSQ 412
QY 415 IECEKIITPOTWETDEHYV---KGATFSLSHKFSQMLYWEPHNRFEELANCYLVGGGTHP 471
DB 413 LVQRIFTRO---TSRHAWTALIGLSLFIETPPSLTQGLFPAANATRH---SNLYLVAAGTHP 466
QY 472 GSGLPITYESARISAKLI 489
DB 467 GAGIPGVVGLAESTASLM 484

RESULT 15

US-08-096-623A-10
Sequence 10, Application US/08096623A
Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H.
APPLICANT: Yarger, James G.
APPLICANT: Yen, Hwei-Che B.
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid

Search completed: February 29, 2004, 14:54:51
Job time : 18.5283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

% protein - protein search, using sw model

run on: February 29, 2004, 14:26:38 ; Search time 60.417 Seconds
(without alignments)
2389.754 Million cell updates/sec

file: US-09-941-947A-22

effect score: 2711
sequence: 1 WANTKHIIIVGAGPGGLCAG.....KRVRFKDIASHAWLKAKA 511

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atadase : A Genesep 29Jan04: *
1: genesep1980s: *
2: genesep1990s: *
3: genesep2000s: *
4: genesep2001s: *
5: genesep2002s: *
6: genesep2003as: *
7: genesep2003bs: *
8: genesep2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2711	100.0	511	5 AAE22309	Methylomo
2	2711	100.0	511	5 ABG61588	High grow
3	2711	100.0	511	5 AAU80332	Methylomo
4	2711	100.0	511	6 ADA14534	Methylomo
5	929.5	34.3	490	5 ABA49224	Listeria
6	929.5	34.3	490	6 ABU32531	Protein e
7	783.5	28.9	499	2 AAR64269	Anabaena
8	713	26.3	494	6 ABM70123	Phototrab
9	703.5	25.9	492	5 AAE22314	Pantoea s
10	703.5	25.9	492	6 AA016021	Pantoea s
11	703.5	25.9	492	6 ABP96688	Pantoea s
12	702.5	25.9	502	6 ABM72613	Staphyloc
13	702.5	25.9	502	6 ADA14542	Staphyloc
14	700.5	25.8	492	2 AAR07466	Polyepti
15	700.5	25.8	492	2 AAW82257	C. utilis
16	700.5	25.8	492	2 AAW87889	Protein e
17	700.5	25.8	492	2 AAU99099	Erwinia u
18	700.5	25.8	492	2 AAY26333	Erwinia u
19	673	24.8	582	2 AAW22499	Blaffia d
20	651	24.0	582	6 ABP97465	Blakeslea
21	640.5	23.6	494	2 AAU00871	Flavobact
22	638.5	23.6	494	2 AAW69532	Flavobact
23	624.5	23.0	526	2 AAR95697	Erythroba
24	614	22.6	489	2 AAU01122	Phytoene
25	614	22.6	489	2 AAW00173	Phytoene

26	614	22.6	489	2 AAW32472	Erwinia h
27	611	22.5	489	2 AAR13985	Phytoene
28	611	22.5	489	2 AAR13984	Phytoene
29	611	22.5	489	2 AAW01123	Phytoene
30	611	22.5	489	2 AAW00174	Phytoene
31	611	22.5	489	2 AAW32475	Erwinia h
32	607.5	22.4	497	5 AAE22310	Methylomo
33	607.5	22.4	497	5 ABG61589	High grow
34	607.5	22.4	497	5 AAU80333	Methylomo
35	607.5	22.4	497	6 ADA14538	Methylomo
36	559.5	20.6	502	6 ABU43877	Protein e
37	502.5	18.5	497	6 ABU16031	Protein e
38	500.5	18.5	497	6 ADA14544	Staphyloc
39	500.5	18.5	500	6 ABM72616	Staphyloc
40	483	17.8	548	4 AAB85728	Enzyme in
41	482.5	17.8	544	5 AA015518	Agropyces
42	480	17.7	548	4 AAB76640	Corynebac
43	480	17.7	548	4 AAB76641	Corynebac
44	480	17.7	548	4 AAG90440	C glutami
45	447.5	16.5	439	4 AAU34200	Staphyloc

ALIGNMENTS

RESULT 1
AAE22309
ID AAE22309 standard; protein; 511 AA.

XX AAE22309;

XX AC

XX XX

XX 07-AUG-2003 (revised)

XX DT 25-JUL-2002 (first entry)

XX XX

XX DE Methylomonas 16a sp. diaphytoene dehydrogenase (CrtnI) enzyme.

XX XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;
XX XX anti-oxidant; steroid; flavour; fragrance; electro-optic application;
XX XX aquaculture; enzyme; diaphytoene dehydrogenase; CrtnI.

XX OS Methylomonas sp.

XX XX WO200218617-A2.

XX XX 07-MAR-2002.

XX PF 04-SEP-2001; 2001WO-US027420.

XX XX 01-SEP-2000; 2000US-0229858P.

XX XX 01-SEP-2000; 2000US-0229907P.

XX XX (DUPO) DU PONT DE REMOURS & CO E I.

XX PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX PI Odum JM, Picataggio SK, Rouviere PE;

XX XX WPI; 2002-351711/38.

XX XX N-PSDB; AAD35507.

XX XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by
XX XX using microorganisms having a nucleic acid molecule encoding enzymes in
XX XX the carotenoid biosynthetic pathway and which metabolize single carbon
XX XX substrates.

XX XX Claim 37; Page 127-129; 156pp; English.

XX XX The invention relates to a method for producing carotenoid compounds. The
XX XX method comprises a transformed metabolising host cell, comprising
XX XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
XX XX encoding an enzyme in the carotenoid biosynthetic pathway, under the
XX XX control of regulatory sequences, and contacting the host cell with carbon
XX XX substrate to produce a carotenoid compound. The method is useful for
XX XX producing carotenoid compounds such as anthraxanthin and astaxanthin, by

using microorganism having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates. The carotenoids have potent anti-oxidant properties useful in diet, and aquaculture elements. The carotenoids are also useful as intermediates in the synthesis of steroids, flavours and fragrances and compounds for potential electro-optic applications. The present sequence is *Methylobacterium* 16a sp. diaphotocene dehydrogenase (CrtN1) enzyme used in the invention. (Updated on 07-AUG-2003 to correct OS field.)

Q Sequence 511 AA;

Query Match 100.0%; Score 2711; DB 5; Length 511;
 Best Local Similarity 100.0%; Pred. No. 6.4e-265;
 Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60
 1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60

61 LMKGVLDENFELCERSEDIYLFELPLSPMYRLLYDDRDIFVYSRENWRAELQRFDEGT 120
 61 LMKGVLDENFELCERSEDIYLFELPLSPMYRLLYDDRDIFVYSRENWRAELQRFDEGT 120

121 DGYEQFMEQERKERNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180
 121 DGYEQFMEQERKERNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180

181 KMLAFPCFOSKYLGMSPWECPCALFTMLPYLHEHYGIYHVQGLNRIIAAQAQVIAENGGE 240
 181 KMLAFPCFOSKYLGMSPWECPCALFTMLPYLHEHYGIYHVQGLNRIIAAQAQVIAENGGE 240

241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKXYTPENL 300
 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKXYTPENL 300

301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKDYTNIRNIPDNKLTDDFSFYQNASAS 360
 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKDYTNIRNIPDNKLTDDFSFYQNASAS 360

361 DDSLAPAGKSALYLVPMNPNDGLDQAHQCNVREQVLDTLGARGLSDIRAHISECKI 420
 361 DDSLAPAGKSALYLVPMNPNDGLDQAHQCNVREQVLDTLGARGLSDIRAHISECKI 420

421 ITPQWETDEHYKGTSLSHKESOMLYWPHRFRFELANCYLVGCGTHPGSGLEIYE 480
 421 ITPQWETDEHYKGTSLSHKESOMLYWPHRFRFELANCYLVGCGTHPGSGLEIYE 480

481 SARISAKLISQKRVRFKDIASHAWLKZAKA 511
 481 SARISAKLISQKRVRFKDIASHAWLKZAKA 511

RESULT 2

ABG61588
 ID ABG61588 standard; protein; 511 AA.

XX AC ABG61588;

XX DT 07-AUG-2003 (revised)

XX DT 27-AUG-2002 (first entry)

XX DE High growth methanotrophic bacterial strain polypeptide #38.

XX KW High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme; methane; methanol; Embden-Meyerhof carbon flux pathway; 16S rRNA; pyrophosphate dependent phosphofructokinase; nitrogen-containing compound; ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill; methanol-containing environment; waste water treatment system; isoprenoid; nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.

XX OS *Methylobacterium* sp.

XX WC200202728-A2.

XX PD 14-MAR-2002.
 XX XX 28-AUG-2001; 2001WC-US026827.
 XX XX 01-SEP-2000; 2000US-0229858P.
 XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX PI Koffas M, Odom JM, Schenzle A;
 XX DR WPI: 2002-454358/48.
 XX DR N-PSDB; ABK33267.
 XX PT New high growth methanotrophic bacterial strain, useful for producing single cell proteins, grows on a C1 carbon substrate, and comprises a functional gene encoding in Embden-Meyerhof carbon pathway.
 XX PS Disclosure; Page 150-152; 157pp; English.
 XX CC The invention relates to a high growth methanotrophic bacterial strain, which grows on a C1 carbon substrate e.g. methane and methanol, and comprises a functional Embden-Meyerhof carbon flux pathway comprising a gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S rRNA. The bacterial strain is useful for the production of single cell protein and for the biotransformation of a nitrogen-containing compound, e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the production of a feed product comprising a protein, carbohydrates and a pigment and for reducing oxygen demand, for removing nitrates and nitrites in methane-containing environments such as landfills, waste water treatment systems or anywhere that methane, oxygen and nitrates are present. The bacterial strain of the invention can be used as a denitrifying agent for the conversion of nitrate or nitrite to nitrous oxide with methane or methanol as a carbon source. It is also used in the production of biomass including proteins, carbohydrates and a wide variety of pigments (particularly for isoprenoid pigments for the purpose of generating animal feeds), in production of terpenoid and carotenoid compounds, useful as pigments and as monomers in polymeric materials and in production of exopolysaccharides at high levels. Sequences ABG61551-ABG61590 represent high growth methanotrophic bacterial strain proteins of the invention. (Updated on 07-AUG-2003 to correct OS field.)

Sequence 511 AA;

Query Match 100.0%; Score 2711; DB 5; Length 511;
 Best Local Similarity 100.0%; Pred. No. 6.4e-265;
 Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60
 DB 1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60

QY 61 LMKGVLDENFELCERSEDIYLFELPLSPMYRLLYDDRDIFVYSRENWRAELQRFDEGT 120
 DB 61 LMKGVLDENFELCERSEDIYLFELPLSPMYRLLYDDRDIFVYSRENWRAELQRFDEGT 120

QY 121 DGYEQFMEQERKERNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180
 DB 121 DGYEQFMEQERKERNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180

QY 181 KMLAFPCFOSKYLGMSPWECPCALFTMLPYLHEHYGIYHVQGLNRIIAAQAQVIAENGGE 240
 DB 181 KMLAFPCFOSKYLGMSPWECPCALFTMLPYLHEHYGIYHVQGLNRIIAAQAQVIAENGGE 240

QY 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKXYTPENL 300
 DB 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKXYTPENL 300

QY 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKDYTNIRNIPDNKLTDDFSFYQNASAS 360
 DB 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKDYTNIRNIPDNKLTDDFSFYQNASAS 360

QY 361 DDSLAPAGKSALYLVPMNPNDGLDQAHQCNVREQVLDTLGARGLSDIRAHISECKI 420
 DB 361 DDSLAPAGKSALYLVPMNPNDGLDQAHQCNVREQVLDTLGARGLSDIRAHISECKI 420

361 DDLAPAGKSAALYLVLPMPNNDGLDWAHCHQVREQLDTLGLRLGLSDIRAHIECKI 420
 421 ITPQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEBELANCYLVGGTTHPGSGLPTIYE 480
 421 ITPQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEBELANCYLVGGTTHPGSGLPTIYE 480
 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511
 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511

RESULT 3

AAU80332 standard; protein; 511 AA.

AAU80332;

15-JUL-2002 (first entry)

Methylomonas 16a ORF8 crtN1 protein sequence.

Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;
 keratinoid; pigment; flavour; fragrance; open reading frame 8; ORF8;
 crtN1; crtN copy1; diaphytoene dehydrogenase enzyme.

Methylomonas sp.

WO200202733-A2.

14-MAR-2002.

29-AUG-2001; 2001WO-US026852.

01-SEP-2000; 2000US-02299078.

(DUPO) DU PONT DE NEMOURS & CO E I.

Cheng Q, Koffas M, Norton KC, Odom JM, Picataggio SK;

Rouviere PE, Schenzle A, Tomb J;

WPI; 2002-383051/41.

N-PSDB; ABK50088.

Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
 isolated from Methylomonas 16a, useful for the production of isoprenoid
 compounds.

Claim 4; Page 78-80; 84pp; English.

The present invention relates to a new nucleic acid molecule encoding an
 isoprenoid biosynthetic enzyme isolated from Methylomonas 16a. The
 invention is useful for obtaining a nucleic acid molecule encoding an
 isoprenoid compound biosynthetic enzyme, and for the microbial production
 of isoprenoid compounds. The molecules of the invention are also useful
 for regulating isoprenoid biosynthesis in an organism and for producing
 recombinant organisms for producing various isoprenoid compounds. The
 nucleic acid is also useful for feed additive, for the production of
 keratinoids and their derivatives, isoprenoid intermediates, and as pure
 products useful as pigments, flavours and fragrances. The present amino
 acid sequence represents the Methylomonas 16a open reading frame 8 (ORF8)
 crtN1 ((crtN copy1) copy 1 of diaphytoene dehydrogenase enzyme) protein
 of the invention, as described above

Sequence 511 AA;

Query Match 100.0%; Score 2711; DB 5; Length 511;

Best Local Similarity 100.0%; Pred. No. 6.4e-265;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANTKHIIIVGAGPGCLCAGMLLSQKGFVSIPIFGHAIIGGNRPINMNGFTFTGTFL 60

1 MANTKHIIIVGAGPGCLCAGMLLSQKGFVSIPIFGHAIIGGNRPINMNGFTFTGTFL 60

61 LMKGVLDENFELCHRRSEDYLFPLSPWYLLYYDRDIFVYSORENRAELQRFVDEGT 120
 61 LMKGVLDENFELCHRRSEDYLFPLSPWYLLYYDRDIFVYSORENRAELQRFVDEGT 120
 121 DGYEQFMEQERRKFNALYPCITRDYSSLSKSFSLDLIKALWLAFFPKSVFNNLGOYFNOE 180
 121 DGYEQFMEQERRKFNALYPCITRDYSSLSKSFSLDLIKALWLAFFPKSVFNNLGOYFNOE 180
 181 KRLAPCFQSKYLGMSPEWECPALFTMLPYLEHYGIYHVKGANRIARAQAQVIAENGGE 240
 181 KRLAPCFQSKYLGMSPEWECPALFTMLPYLEHYGIYHVKGANRIARAQAQVIAENGGE 240
 241 IHLNSEIESLIIENGAAKGVKLOHGAELRGDEVIINADFAAMTHLVKPGVLKKTTPENL 300
 241 IHLNSEIESLIIENGAAKGVKLOHGAELRGDEVIINADFAAMTHLVKPGVLKKTTPENL 300
 301 KQREYSCSTFMYLGLDKIYDLPHTTIVPAKDYTNIRNIDNKTLTDDPSFYVQNASAS 360
 301 KQREYSCSTFMYLGLDKIYDLPHTTIVPAKDYTNIRNIDNKTLTDDPSFYVQNASAS 360
 361 DDLAPAGKSAALYLVLPMPNNDGLDWAHCHQVREQLDTLGLRLGLSDIRAHIECKI 420
 361 DDLAPAGKSAALYLVLPMPNNDGLDWAHCHQVREQLDTLGLRLGLSDIRAHIECKI 420
 421 ITPQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEBELANCYLVGGTTHPGSGLPTIYE 480
 421 ITPQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEBELANCYLVGGTTHPGSGLPTIYE 480
 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511
 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511

RESULT 4

ADAL4534
 ID ADAL4534 standard; protein; 511 AA.

ADAL4534;

06-NOV-2003 (first entry)

Methylomonas sp. 16a CrtN protein SEQ ID NO:4.

carotenoid biosynthetic enzyme; CBE;

omega-aldehyde-functionalised carotenoid;

omega-carboxy-functionalised carotenoid; C30 carotenoid;

carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;

CrtN.

Methylomonas sp.

WO2003069917-A2.

21-AUG-2003.

11-FEB-2003; 2003WO-US004150.

11-FEB-2002; 2002US-0355939P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Cheng Q, Norton KC, Tao L;

WPI; 2003-697524/66.

N-PSDB; ADAL4533.

New nucleic acid encoding carotenoid biosynthesis enzymes, useful for
 preparing functionalized carotenoids, e.g. pigments and antioxidants,
 also encoded polypeptides.

Example 4; Page 91-93; 125pp; English.

CC The present invention describes nucleic acid sequences encoding
 CC carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE
 CC polypeptides; (2) a chimeric gene (CG) containing a CBE linked to
 CC regulatory sequences; (3) host cell transformed with CG; (4) obtaining a
 CC nucleic acid that encodes a CBE; (5) products of method (4); (6)
 CC producing an omega-aldehyde-functionalised carotenoid (X); (7) producing
 CC an omega-carboxy-functionalised carotenoid (Y); (8) producing C30
 CC carotenoids (Z); and (9) regulating carotenoid biosynthesis in an
 CC organism. CBE sequences have antioxidant activity. CBEs can be used for
 CC microbial production of aldehyde- or carboxy-functionalised carotenoids
 CC and C30 carotenoids, and are potentially useful as pigments and
 CC antioxidants, e.g. as feed additives. The present sequence represents a
 CC CrtN CBE protein isolated from *Methylobacterium* sp. strain 16a, which is used
 CC in the exemplification of the present invention.
 CC
 CC Sequence 511 AA:
 CC
 CC Query Match 100.0%; Score 2711; DB 6; Length 511;
 CC Best Local Similarity 100.0%; Pred. No. 6.4e-265;
 CC Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 2Y 1 MANTGHHIIIVGAGPGGLCAGMLLSORGFKVSIFDKHAEIGGRNRPINMGFTDGTPTL 60
 CC 2b 1 MANTGHHIIIVGAGPGGLCAGMLLSORGFKVSIFDKHAEIGGRNRPINMGFTDGTPTL 60
 CC 2Y 61 LMKGVLDLDFELCRRSEDIYLFPLSPMYRLLYDDRDIFVYSDRENRAELQRFDEGT 120
 CC 2b 61 LMKGVLDLDFELCRRSEDIYLFPLSPMYRLLYDDRDIFVYSDRENRAELQRFDEGT 120
 CC 2Y 121 DGTQFMQEKRRFNALYPCITRDYSSLSKFLSLDLIKALPWLAPFKSVNNIGQYFNOE 180
 CC 2b 121 DGTQFMQEKRRFNALYPCITRDYSSLSKFLSLDLIKALPWLAPFKSVNNIGQYFNOE 180
 CC 2Y 181 KRLAFQCSQKVLGNSPWECPALFTMLPYLHEHYGIHVKAGLNRLAAMAQVIAENGGE 240
 CC 2b 181 KRLAFQCSQKVLGNSPWECPALFTMLPYLHEHYGIHVKAGLNRLAAMAQVIAENGGE 240
 CC 2Y 241 IHLNSESIESLIENGAAGVKLQHGAEIRGDEVIIINADPAHAMTHLVKPVGLKXYTPENL 300
 CC 2b 241 IHLNSESIESLIENGAAGVKLQHGAEIRGDEVIIINADPAHAMTHLVKPVGLKXYTPENL 300
 CC 2Y 301 KOREVSCSTFMYLGLDKIKYDLPHHTIVFAKDYTNIRNIFDNKTLIDDFSFFVQNASAS 360
 CC 2b 301 KOREVSCSTFMYLGLDKIKYDLPHHTIVFAKDYTNIRNIFDNKTLIDDFSFFVQNASAS 360
 CC 2Y 361 DDLAPACKSALYLVPMPPNDSGLDMQAHQCNVREQVLDTLGARGLSLDIRAHIECEKI 420
 CC 2b 361 DDLAPACKSALYLVPMPPNDSGLDMQAHQCNVREQVLDTLGARGLSLDIRAHIECEKI 420
 CC 2Y 421 ITPQWETDEHYKCATFSLSHKFSQMLYWRPHNRPFEELANCYLVGGTHPGSGLPTIYE 480
 CC 2b 421 ITPQWETDEHYKCATFSLSHKFSQMLYWRPHNRPFEELANCYLVGGTHPGSGLPTIYE 480
 CC 2Y 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511
 CC 2b 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511
 CC
 CC RESULT 5
 CC ABB49224
 CC ID ABB49224 standard; protein; 490 AA.
 CC XX
 CC AC ABB49224;
 CC XX
 CC XX 05-FEB-2002 (first entry)
 CC XX
 CC DE *Listeria* monocytes protein #1928.
 CC XX
 CC XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 CC XX vitamin B12; bacterial infection; disease.
 CC XX
 CC OS *Listeria* monocytes.
 CC XX

PN WO200177335-A2.
 XX 18-OCT-2001.
 XX 11-APR-2001; 2001WO-FR001118.
 XX 11-APR-2000; 2000FR-00004629.
 XX (INSP) INST PASTEUR.
 PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Deboux P;
 PI Dusurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cosset P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX WPI; 2002-010914/01.
 XX
 XX Genomic sequence for *Listeria* monocytes, useful e.g. for treatment
 XX and prevention of *Listeria* and related bacterial infections, and related
 XX polypeptides.
 XX
 XX Claim 6; SEQ ID NO 1929; 192pp; French.
 XX
 XX The present invention relates to the genome sequence of *Listeria*
 XX monocytes EGD-e (see ABA03041). The genome sequence and fragments of
 XX it are useful for selecting probes and primers for detecting genes in *L.*
 XX monocytes and related organisms, and for studying genetic
 XX polymorphisms and other genomes. The present sequence is a protein
 XX encoded by the genome sequence of the present invention. Proteins
 XX expressed from the genome sequence are useful for raising specific
 XX antibodies, identification of *L.* monocytes and related organisms, and
 XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 XX B12. The genome sequence and proteins encoded by it are also useful for
 XX selecting compounds that regulate gene expression and cell replication
 XX and modulate *L.* monocytes-related diseases. In addition, the genome
 XX sequence and proteins encoded by it are useful in pharmaceutical and
 XX vaccine compositions for the treatment or prevention of infections by *L.*
 XX monocytes and related organisms. Note: The sequence data for this
 XX patent did not form part of the printed specification, but was obtained
 XX in electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 490 AA;
 XX
 XX Query Match 34.3%; Score 929.5; DB 5; Length 490;
 XX Best Local Similarity 37.6%; Pred. No. 1.5e-84;
 XX Matches 184; Conservative 98; Mismatches 199; Indels 9; Gaps 1;
 XX
 XX 5 KHIIVGAGPGGLCAGMLLSORGFKVSIFDKHAEIGGRNRPINMGFTDGTPTL 64
 XX 7 KRIATIGAGPGGLCAGMLLSORGFKVSIFDKHAEIGGRNRPINMGFTDGTPTL 66
 XX 65 VLDEMFCRRSEDIYLFPLSPMYRLLYDDRDIFVYSDRENRAELQRFDEGT 124
 XX 67 VLTSFMDNENILYVSLIPNHTLYPKDITFPLYSQSETKAVITQYFPCEDGFD 126
 XX 125 QWQERKRRFNALYPCITRDYSSLSKFLSLDLIKALPWLAPFKSVNNIGQYFNOE 184
 XX 127 RPKNTKMLYISPLNQFNYSILDFRPTTLTAIPSLMGRSLMDLISYFNSKILRL 186
 XX 185 APCFOSKYLGMSPWECPALFTMLPYLHEHYGIHVKAGLNRLAAMAQVIAENGGE 244
 XX 187 AFSLQVRVILGSPWIDIPAAYSIIPSEYVYGFTHFGICQNKIVAMQVVTENKGFEN 246
 XX 245 SPESLIENGAAGVKLQHGAEIRGDEVIIINADPAHAMTHLVKPVGLKXYTPENL 304
 XX 247 SVTSEFSGKEITGAVLANGKTIADYITFLNDFIVSLTN-----EHPDKLDTKE 297
 XX 305 YSCSTFMYLGLDKIKYDLPHHTIVFAKDYTNIRNIFDNKTLIDDFSFFVQNASAS 364

```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
298 YSSAFILYGLTKVLPFSHQSIIPENYREFAHNTMHKILSKDIAIHLNPSATDNTM 357
365 APAGKSALVLPMPNNDGLDQAHQCNVRQVLDITLGAHGLSDIRAHIECEKIITPQ 424
358 APINSSIRIMVPVNNNTSNIDKKETAPFRQLVLETWKERLEIPDLSESQIEEYIITPI 417
425 TWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEELANCYLVGGGTHPGSGLPIYESARI 484
418 DWEKHYVHKGAFGLQHLWRQHGVLHPSKSPKPNLFVIGAGAMSGSLPFIENAIQI 477
485 SAKLISQKX 494
478 ATQKFLQXK 487

```

RESULT 6

B032531
X ABU32531 standard; protein; 490 AA.

X ABU32531;

X 19-JUN-2003 (first entry)

X Protein encoded by Prokaryotic essential gene #18058.

X Antisense; prokaryotic essential gene; cell proliferation; drug design.

X Listeria monocytogenes.

X W0200277183-A2.

X 03-OCT-2002.

X 21-MAR-2002; 2002WO-US009107.

X 21-MAR-2001; 2001US-00815242.

X 06-SEP-2001; 2001US-00948993.

X 25-OCT-2001; 2001US-0342923P.

X 08-FEB-2002; 2002US-00072851.

X 06-MAR-2002; 2002US-0362699P.

X (ELIT-) ELITRA PHARM INC.

X Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW,

X Wall D, T-awick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

X WPI; 2003-029926/02.

X R-PSDB; ACA36401.

X Claim 25; SEQ ID NO 60455; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 490 AA;

Query Match 34.3%; Score 929.5; DB 6; Length 490;

Best Local Similarity 37.6%; Pred. No. 1.5e-84;

Matches 184; Conservative 98; Mismatches 199; Indels 9; Gaps 1;

```

QY 5 KHIIVGAGPGGLCAGMLLSQSGFKVSIYFDKHAETGGNRINNGFTDGTGPTLLMKG 64
DB 7 KKIATGAGPGGLAAGMLLSQSGYQNTYERKDRIGRTAHRMGKYSFDVGPSTMTM 66
QY 65 VLDEMFELCERSESDYLEPLSPMYRLYDDRDIFVYSDRENMAELQRFDEGTGYE 124
DB 67 VLTSLFDCNRNILDVSVLLPINPIHTLYPKDITFPLYSYD(SETKAVIQTYPPGEDGFD 126
QY 125 QNMQQRKFNALYPCITRDYSSLSKSLDILKALPWLAAPKSVFNNLQYFNQKKRL 184
DB 127 RPKMKNKMLYISPLNQFNYSLSDFRPTTIRAPSLTGRSLMDLRSYFNSKYLRL 186
QY 185 APCFSQKILGMSPECPALFTMLPYLEHYHYHVGGLNEIAAAMAQVIAENGGEIHLN 244
DB 187 AFSLOWYILGMSFWDIPRAYSIIPSEYTYGTFHPIGGOKIIVEMQQVYVTKNKKFTFN 246
QY 245 SEIESLIENGAAKGVKLOHGAELRGDEVIINADFAHAMTHAVKPGVLKKYTPENLKORE 304
DB 247 SEVTEPESNGKEITGAVLANGKTIADYVFTNLDFTYSLTN-----EHPDKLDTKE 297
QY 305 YSCSTPMYVLGDKLYLDLPHHTIVFAKDYTTINIRIFDKLTDDPSFYVQNASASDLSL 364
DB 298 YSSAFILYGLTKVLPFSHQSIIPENYREFAHNTMHKILSKDIAIHLNPSATDNTM 357
QY 365 APAGKSALVLPMPNNDGLDQAHQCNVRQVLDITLGAHGLSDIRAHIECEKIITPQ 424
DB 358 APINSSIRIMVPVNNNTSNIDKKETAPFRQLVLETWKERLEIPDLSESQIEEYIITPI 417
QY 425 TWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEELANCYLVGGGTHPGSGLPIYESARI 484
DB 418 DWEKHYVHKGAFGLQHLWRQHGVLHPSKSPKPNLFVIGAGAMSGSLPFIENAIQI 477
QY 485 SAKLISQKX 494
DB 478 ATQKFLQXK 487

```

RESULT 7

AAR64269

ID AAR64269 standard; protein; 499 AA.

XX AAR64269;

AC AAR64269;

XX 16-OCT-2003 (revised)

DT 27-AUG-2003 (revised)

DT 02-OCT-1995 (first entry)

XX Anabaena epsilon-carotene desaturase protein.

XX Anabaena, epsilon-carotene desaturase gene; SAN380H; J852; plant;

KW phytoene desaturase inhibiting herbicide; norflurazon; fluridone;

KW Brwinia uredovora; lycopene converting activity; transit peptide.

DS Nostoc sp: "PCC 7120".
XX Key Location/Qualifiers
XX Misc-difference 296 /note= "encoded by ATC"
XX Misc-difference 299 /note= "encoded by TCA"
XX Misc-difference 300 /note= "encoded by GAC"
XX Misc-difference 302 /note= "encoded by CTT"
XX Misc-difference 303 /note= "encoded by TAT"
XX Misc-difference 304 /note= "encoded by TTG"
XX Misc-difference 305 /note= "encoded by GGT"
XX Misc-difference 306 /note= "encoded by ATC"
XX Misc-difference 307 /note= "encoded by AAT"
XX Misc-difference 308 /note= "encoded by CGC"
XX Misc-difference 309 /note= "encoded by CGC"
XX Misc-difference 310 /note= "encoded by TAC"
XX Misc-difference 311 /note= "encoded by GAA"
XX Misc-difference 312 /note= "encoded by GAT"
XX Misc-difference 313 /note= "encoded by TTA"
XX Misc-difference 314 /note= "encoded by CCT"
XX Misc-difference 315 /note= "encoded by CAT"
XX Misc-difference 316 /note= "encoded by CAT"
XX Misc-difference 317 /note= "encoded by CAA"

JP06343473-A.

20-DEC-1994.

08-JUN-1993; 93JP-00163926.

08-JUN-1993; 93JP-00163926.

(KIRI) KIRIN BREWERY KK.

WPI: 1995-069305/10.

N-PSDB; AAQ80736.

Prod'n of bleaching herbicide-resistant plants - by transformation with the Erwinia phytoene desaturase gene.

Example 9; Page 11-13; 20pp; Japanese.

The sequence of the Anabaena strain PCC7120 epsilon-carotene desaturase protein. The corresponding gene was used in a method for the production of plants resistant to epsilon-carotene desaturase inhibiting herbicides e.g. SAN380H, J852, etc., or phytoene desaturase inhibiting herbicides e.g. norflurazon, fluridone, etc., by introducing into the plant a DNA sequence that encodes a polypeptide, preferably the Erwinia-derived phytoene desaturase e.g. the E.uredovora crtI gene, or a variant e.g. the Anabaena sequence given in the specification, having a phytoene to lycopene converting activity. The DNA sequence also contains a sequence coding for a transit peptide. (Updated on 27-AUG-2003 to correct OS field.) (Updated on 16-OCT-2003 to standardise OS field)

SQ Sequence 499 AA;
Query Match 28.9%; Score 783.5; DB 2; Length 499;
Best Local Similarity 35.0%; Pred. No. 9.2e-70;
Matches 175; Conservative 103; Mismatches 201; Indels 21; Caps 10;
QY 4 TKHIIYAGGGGICAGMLLSQRFKYSIFDKHAEITGGRRPINMNGFTDGTGPTLLMK 63
DB 2 SKKVAIVGAGGGGATATRLAGLQIQVIEFAARVGGRRGGEVSVYADTGTTLQLP 61
QY 64 GVLDEMPCLCERSELYLEFLPLSPMYRL-LYDNRDIFVYSRENMRALQRFDECTDG 122
DB 62 HLYKELFEELGNFADYVQLKRLPEYTRLAKFMDGTOLDITDLQSPKTLATLRSDLPLA 121
QY 123 YEQFMEQERKFNALY-PCITRDYSSLSKLSL-DLIKAL---PWLAFPKSVNNLQGYF 177
DB 122 FDRWYSBHIRKYEGLYKPYLAGPARSIFGYLRPDDLKMLSPFPW----ENLQHFWRFP 177
QY 178 NOEKLAFQFQSKYLQMSWPCFALFTMLPYLEHEXYGIYHVKGGLNRIAAAOVIAEN 237
DB 178 QDERLVYDLRYPSKYLGMHPTVASSVSLIPFLEFSQGVNKPVGGFALAAQGLANAQDL 237
QY 238 GGBHLNSETBSLIIENGAAGKYLQHGAEIRGDEVILINADFAHMTLVKPGVLKKTYP 297
DB 238 GVKIHLSPVHQIWIIDQGVGRGLELADASHQFDTVVINADFAFAYVHLLPTTSARGYTD 297
QY 298 ENLKQREYSCSTFMLYGLDKIY-DLPHRTIVFAKDYTTIRNI----FENKLTLD-DPS 351
DB 298 NKLGMQFSCSTFMLYGLINRYEDLPHQIYL---SDMIRLERPWVDDSDALDETDP 353
QY 352 FYVONASASDOSLAPAGKSAIYLVPMNDSDGLDQWQACQNYREQVLDLTGLARGLSDI 411
DB 354 FYVCNFTIIDPSNAPAGHSTLFLVLPITNTSYAVDWDIKQKSYTDFILKRLHL-LGVHNI 412
QY 412 RAHIBCEKIIIPQWETDEHVYKGAFTSLGHKFSQMLYWEPHNRFEELANCYLVGCGTHP 471
DB 413 EGHIVTQSCYTAQSNLDDYRVHLGAVFNLSHNLTLQLGPPRPPIRSENIAGLYWIGGAVHP 472
QY 472 GSGLPITYESARISAKLISQ 491
DB 473 GSGLLTILEASRSAAAGFIHQ 492
RESULT 8
ABW70123
ID ABW70123 standard; protein; 494 AA.
XX
AC ABW70123;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #3220.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN W0200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002NO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;

XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX
XX Claim 2; SEQ ID NO 3220; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens proteins
XX
XX Sequence 494 AA;

Query Match 26.3%; Score 713; DB 6; Length 494;
Best Local Similarity 31.4%; Pred. No. 1.2e-62;
Matches 159; Conservative 109; Mismatches 212; Indels 24; Gaps 10;
Y 8 IIVGAPGGICAGMLLSQKQKVSIFDKHAEIIGGRNRPINMGFTFDGPTFLMKGVLD 67
b 5 LVTCAGFGGLALAIRLQAGIPTCEILEQDKPGRAYVYKEQGFDFDAGPTVITAPNVE 64
Y 68 EMPELCERSEDEYLEPLPSPMYRLYDDRDIFVY--SDRENRAELQRFVDEGTGYQF 126
b 65 ELTQAGKRWADYVDLLPVHPFVRLCWESKGFVDNDQOHLAQIHTFNPRDVGNYRFP 124
Y 127 MEQERKRNALYPCI--TRYSSLSKFLSL--DLIKALPWLAPPKSVFNNLQGYFNQEKMR 183
b 125 LDYSRAFNQEGYKLGCTVPLSPFDMLSAAPQLIRLHAM---RSVYQVAFIKDESILR 180
Y 184 LAFCSQKYLGSWPCPALFTMLPYLEHEGYIYHVGGLNRIAMAQVIAENGGEIHL 243
b 181 QAFSFLSLIGGNPFAASIIYTHIALRENGWVFPFGTSSALVEAMVKLFTDGGIEEL 240
Y 244 NSBIESLIIENGAAGKGVKLQHGAEIAGDEVIINADPAHAMTHLV---KPGVLKXYTPENL 300
b 241 NAKVKFTTHGNRTVCGVLADGENMTCDVVASNADVIHTYKHLGCHPVGIAART---L 297
Y 301 KQREYSCSTFMGLYGLDKYI-DLPHHTIVFAKDYTNIRNIPDNKTLTDDTSFYVQNAS 359
b 298 VRKMSNSFLVYFGLNHHHTQLAHTVCFGPYKELIEDIFDHLRSEDPSLYLHAPS 357
Y 360 SDDSLAPAGKSAIYLVLPWPN--NDSGLDQACQNVREOVLDTLGLRGLSLDIRAHICE 418
b 358 TDSLAPRGCAIYLVLPWPHLGTANLWIDEGPRIDRIFAYL--EKYMPGLLKLQVYH 416
Y 419 KIITPQWTDEHYKGAFTSLSHKFSQMLYWRPNRNFEELANCYVGGGTHPGSLPTI 478
b 417 RIPTPFDQDLNAHLGSAFSPPELLTQSAWEPFNHNRNDNLNLYLVAGTHPGAGIPGV 476
Y 479 YESARISAKLISKQHVRFKDLA 501
b 477 IGSKATATIM-----LEDIA 492

RESULT 9
AAE22314
ID AAE22314 standard; protein; 492 AA.
XX AC AAE22314;
XX DT 25-JUL-2002 (first entry)
XX DE Pantoea stewartii phytoene desaturase (CrtL) enzyme.
XX KW Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;
XX anti-oxidant; steroid; flavour; fragrance; electro-optic application;
XX aquaculture; enzyme; phytoene desaturase; CrtL.
XX OS Pantoea stewartii.
XX WO200218617-A2.
XX PD 07-MAR-2002.
XX PF 04-SEP-2001; 2001WO-US027420.
XX PR 01-SEP-2000; 2000US-0229858P.
XX PA 01-SEP-2000; 2000US-0229907P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;
XX PI Odom JM, Picataggio SK, Rouviere PE;
XX DR WPI; 2002-351711/38.
XX DR N-PSDE; AAD35512.
XX PT Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by
XX using microorganisms having a nucleic acid molecule encoding enzymes in
XX the carotenoid biosynthetic pathway and which metabolize single carbon
XX substrates.
XX PS Claim 19; Page 141-143; 156pp; English.
XX CC The invention relates to a method for producing carotenoid compounds. The
XX method comprises a transformed metabolising host cell, comprising
XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule
XX encoding an enzyme in the carotenoid biosynthetic pathway, under the
XX control of regulatory sequences, and contacting the host cell with carbon
XX substrate to produce a carotenoid compound. The method is useful for
XX producing carotenoid compounds such as anthraxanthin and astaxanthin, by
XX using microorganisms having a nucleic acid molecule encoding enzymes in
XX the carotenoid biosynthetic pathway and which metabolize single carbon
XX substrates. The carotenoids have potent anti-oxidant properties useful in
XX diet, and aquaculture elements. The carotenoids are also useful as
XX intermediates in the synthesis of steroids, flavours and fragrances and
XX compounds for potential electro-optic applications. The present sequence
XX is Pantoea stewartii phytoene desaturase (CrtL) enzyme used in the
XX invention
XX SQ Sequence 492 AA;
Query Match 25.9%; Score 703.5; DB 5; Length 492;
Best Local Similarity 31.7%; Pred. No. 1.1e-61;
Matches 159; Conservative 105; Mismatches 210; Indels 25; Gaps 10;
QY 5 KHIIIVGAPGGICAGMLLSQKQKVSIFDKHAEIIGGRNRPINMGFTFDGPTFLMKG 64
Db 2 KPTTVTGAGFGGLALAIRLQAGIPTCEILEQDKPGRAYVYKEQGFDFDAGPTVITDPS 61
QY 65 VLDEMPCLERSEDEYLEPLPSPMYRLYDDRDIFVY--SDRENRAELQRFVDEGTGY 123
Db 62 AIEELPALAGKQLKQYVELLPVTPFVRLCWESKGFVDNDQOHLAQIHTFNPRDVG 121
QY 124 EQMEQERKRNALYPCIITRDYSSLK---SPISL-DLIKALPWL---FPKSVFNNLQ 175
Db 122 RAFLDYSRAVFNQEGY-----LKGTVPLSPFDMLSAAPQLIRLHAMAPOLAKLQAMRSVYSKVAG 172

QY 176 YNQKMLAFCQSKYLGMSWPCALFTMLPYLEHSYGIYHVKGGLNRIAAQAQVIA 235
 DB 173 YIEDEHLQAQSFHSLVGGNPFATSSYITLIHALEREWGVPFRGGTGVAGLVNMIKLFQ 232
 QY 236 ENGGEIHLNIESLIIENGAAKGVKLOHGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
 DB 233 DLGGEVVLNARVSHMETVGDKIQAQVLEDGRFRFTCAVASNADVVHTYRDLSSQHPAAK 292
 QY 294 KYTPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVPAKYTTNIRNIFDNKTLTDDDFSF 352
 DB 293 Q--AKKQSKMSNSLFVLYFGLNHHHDQLAHTVCGPRYRELIHEIFNHDGLAEDFSL 350
 QY 353 YVONASASDDSLAPACKGALYLVMPMN-NDSGLDQWQHCONVREQVLDITLGLRLGSDI 411
 DB 351 YLHAPCVTDPSLAPGCGSYVYLAPEVHLGTANLDWAVEGPRLDRIYDYLEQHY-MPGL 409
 QY 412 RAHIECEKIITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPFEELANCYLVGCGTHP 471
 DB 410 RSQLVTHRMFTFFDRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
 QY 472 QSGLPFTIVESARISAKLI 489
 DB 470 GAGIPGVIGSAKATAGLM 487

RESULT 10

QY AAO16021 standard; protein; 492 AA.

DB AAO16021;

QY 20-FEB-2003 (first entry)

DB Pantoea stewartii phytoene desaturase.

QY Carotenoid; crt.

DB Pantoea stewartii.

QY WO200279395-A2.

DB 10-OCT-2002.

QY 25-JAN-2002; 2002WO-US002124.

DB 26-JAN-2001; 2001US-0264329P.

QY 04-MAY-2001; 2001US-0288984P.

DB (CRGI) CARGILL INC.

QY De Souza ML, Kollmann SR, May CA, Schroeder WA;

DB WPI: 2003-075455/07.

QY N-PSDB; ABT14193.

DB Novel isolated nucleic acid useful e.g. to engineer host cells with the

QY ability to produce particular carotenoids and polypeptides useful in cell

DB -free systems to make particular carotenoids.

QY Claim 26; Page 63-64; 74pp; English.

DB The invention comprises the amino acid and coding sequence of a number of

QY carotenoid (crt)-related proteins. The crt-related DNA and protein

DB sequences of the invention are useful for engineering cells which are

QY able to produce carotenoids. The present amino acid sequence represents a

DB crt-related protein of the invention

QY Sequence 492 AA;

DB Query Match 25.9%; Score 703.5; DB 6; Length 492;

QY Best Local Similarity 31.7%; Pred. No. 1.1e-61;

DB Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 5 KHIITVAGPGLCAGMLLSQSGPKVSIQFKHATIGQNRNINNGFTFDGPTFLLMKG 64
 DB 2 KPTTIVAGPGLALAIRQAAGIPVLLLEQDRPGGRVAVYQPGGTFDAGPTVITDPS 61
 QY 65 VLDEMFELCERSESDYELPLSPMYRLLYDDRIDFVY-SDRENMAELORVDEGTDGY 123
 DB 62 AIEELFALAGKQLKDYVELLPVTFYRLCWESGKVFYNDADQAOLEAQIQGFNEDVAGY 121
 QY 124 EOPMEQERKFNALYPCITRDYSSLK-----SFLSL-DLIKALPWLA---FPKSVFNNGQ 175
 DB 122 RAPLYSRAVENEGY-----LKLGTVPFLSFKXDMLEAAPOLAKLQAWRSVYSKVAG 172
 QY 176 YNQKMLAFCQSKYLGMSWPCALFTMLPYLEHSYGIYHVKGGLNRIAAQAQVIA 235
 DB 173 YIEDEHLQAQSFHSLVGGNPFATSSYITLIHALEREWGVPFRGGTGVAGLVNMIKLFQ 232
 QY 236 ENGGEIHLNIESLIIENGAAKGVKLOHGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
 DB 233 DLGGEVVLNARVSHMETVGDKIQAQVLEDGRFRFTCAVASNADVVHTYRDLSSQHPAAK 292
 QY 294 KYTPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVPAKYTTNIRNIFDNKTLTDDDFSF 352
 DB 293 Q--AKKQSKMSNSLFVLYFGLNHHHDQLAHTVCGPRYRELIHEIFNHDGLAEDFSL 350
 QY 353 YVONASASDDSLAPACKGALYLVMPMN-NDSGLDQWQHCONVREQVLDITLGLRLGSDI 411
 DB 351 YLHAPCVTDPSLAPGCGSYVYLAPEVHLGTANLDWAVEGPRLDRIYDYLEQHY-MPGL 409
 QY 412 RAHIECEKIITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPFEELANCYLVGCGTHP 471
 DB 410 RSQLVTHRMFTFFDRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
 QY 472 QSGLPFTIVESARISAKLI 489
 DB 470 GAGIPGVIGSAKATAGLM 487

RESULT 11

QY ABP96688

DB ID ABP96688 standard; protein; 492 AA.

QY AC ABP96688;

DB 03-JUN-2003 (first entry)

QY Pantoea stewartii phytoene desaturase SEQ ID NO:6.

DB Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;

QY crtI; crtB; crtZ; phytoene desaturase; enzyme; phytoene; carotenoid.

DB Pantoea stewartii.

QY WO2003016503-A2.

DB 27-FEB-2003.

QY 15-AUG-2002; 2002WO-US026647.

DB 15-AUG-2001; 2001US-0312646P.

QY (DUPO) DU PONT DE NEMOURS & CO E I.

DB Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PS;

QY WPI; 2003-268323/26.

DB N-PSDB; ACC44762.

QY Novel nucleic acid molecule isolated from Pantoea stewartii encoding a

DB carotenoid biosynthetic enzyme, useful for regulating carotenoid

QY biosynthesis in an organism.

DB Claim 4; Page 63-65; 68pp; English.

X The present invention describes Pantoea stewartii carotenoid biosynthetic
 C enzymes (I). More specifically described are the geranylgeranyl
 C pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
 C lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
 C (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to
 C ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating
 C carotenoid biosynthesis in an organism, by over-expressing (I) in an
 C organism, such that the carotenoid biosynthesis is altered in the
 C organism. (I) and the genes encoding (I) are useful for converting
 C phytoene to the carotenoids, for creating recombinant organisms that have
 C the ability to produce various carotenoid compounds, and also for
 C enhancing or manipulating carotenoid compounds. (I) can also be used for
 C producing gene products having enhanced or altered activity

X Q Sequence 492 AA;

Query Match 25.9%; Score 703.5; DB 6; Length 492;
 Best Local Similarity 31.7%; Pred. No. 1.1e-61;
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

Y 5 XHIIIVGAGFGGLCAGMLLSQGFVKVSIFFDKHAEIGGRNRPINMNGFTDGTGPTFLMKG 64
 b 2 KPTVIGAGFGGLAIRLOAGIPVLLLEQDKGGRAYVYQEGFTDAGTVIDPS 61
 Y 65 VLDEMFELCERSEDEYLFPLSPMYRLLYDDRDIPVY-SDENNRAELQRFVDEGTGY 123
 b 62 AIEELPALAGKQLKDYVELLPVTPFYRLCWESGKVFYNDQAOLEAQIQFNPDRVAGY 121
 Y 124 EOPMEQERKFNALYPCITRDYSSLK---SFLSL-DLIKALPMLA---FPKSVFNILGQ 175
 b 122 RAPLDYSRAVFNEY-----LKLGVFPLSFQDMLRAAPQLAKLQWRSVYSKVG 172
 Y 176 YFNOERKRLAFQFSQKYLQSPWECPCALFTMLPYLEHGYIYHVKGSLNRJAAAMQVIA 235
 b 173 YIDELHQAQFPHSLVGNPFATSSIVTLIHALERWGVNPPRGGTGALVNGMIKLFQ 232
 Y 236 ENGCEHIANSEIESLIENGAAGVKLOHGAELRGDEVLIINADFAHATHLVK--PGVLK 293
 b 233 DLUGGVNARVSHMETVGDKIQAQVLEDRGFETFCAVASNADVVTYRDLISQHPAAK 292
 Y 294 KVTPEMLKOREYSCSTFMYLGLDKIYD-LPHHTVFAKDYTNIRNIFDNKLTDDFSF 352
 b 293 Q--AKKLQSKRNSLSFLVYFGLNHHHQLAHETVCFGRYELIHEIFNHDGLAEDPSL 350
 Y 353 YVONASASDSDSLAPAGKSALYVLPMPN-NDSGLDQAHQCNVREQVLDLTGARLGLSDI 411
 b 351 YLHAPCVTDPSLAPEGCGSYVYVLPVPHLGTANLDWAVEGPELRDRIIDYLEQHY-MEGL 409
 Y 412 RAHIECEKIITPQTWETDEHYKGFATSLSHKFSQMLYWRPNRFEELANCVLVGGGTHP 471
 b 410 RSQLVTHRMFTPDFDELANQGSASFSEPILTQSAMFRPNRDKHIDNLYLVGAGTHP 469
 Y 472 GSGLPITYSARISAKLI 489
 b 470 GAGIPGVIGSAKATAGLM 487

RESULT 12

BM72613
 D ABM72613 standard; protein; 502 AA.

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

X X

PN WC200294868-A2.

PD 28-NOV-2002.

XX 27-MAR-2002; 2002NO-IB002637.

PF 27-MAR-2001; 2001GB-00007661.

PR (CHIR-) CHIRON SPA.

XX Masignani V, Mora M, Scarselli M;

XX WPI; 2003-120786/11.

DR N-PSDB; AC74173.

XX New Staphylococcus aureus protein, useful as a vaccine for treating or

PT preventing Staphylococcal infection, specifically an infection caused by

PT S. aureus, e.g. sepsis.

XX Claim 1; SEQ ID NO 3706; 49pp; English.

XX The invention relates to novel genes and encoded proteins from

CC Staphylococcus aureus. A composition comprising the S. aureus protein, a

CC nucleic acid encoding the protein, or an antibody to the protein, is

CC useful as a pharmaceutical, particularly as a vaccine for treating or

CC preventing infection due to Staphylococcus bacteria, specifically an

CC infection caused by S. aureus. The composition is particularly useful for

CC treating or preventing sepsis in a patient. The composition can also be

CC used for diagnostics. The protein is also used in an assay for enzymatic

CC studies and as a target for antibiotics. This sequence represents one of

CC the novel S. aureus proteins of the invention

XX Sequence 502 AA;

Query Match 25.9%; Score 702.5; DB 6; Length 502;

Best Local Similarity 31.5%; Pred. No. 1.5e-61;

Matches 159; Conservative 108; Mismatches 196; Indels 41; Gaps 12;

QY 7 IIVGAGPGLCAGMLLSQGFVKVSIFFDKHAEIGGRNRPINMNGFTDGTGPTFLMKGV 66

DB 3 IAVIGAGVGLAAARIASQGHVETIFEKNNVGGNNQKXKDGFTDGTGPTFLMKGV 62

QY 67 DEMFELCERSEDEYLFPLSPMYRLLYDDRD-IFVYSORENNRAELQRFVDEGTGYEQ 125

DB 63 KDVPYTAGKNYEDYIELRLQRYIYDVYFDHDDRIIVPTDLAELQQLMESIEPGSTHGFMS 122

QY 126 FMEQERKFNALYPCITR-----DYSSLXSPSLDLIKALPWLAPKSVFNIL 173

DB 123 FLTDVYKYE-----IARYFLERTYRKPSDFYNTSLVQGAKLKTL-----NHA 167

QY 174 GQ----YFNOEKRLAFQFSQKYLQSPWECPCALFTMLPYLEHGYIYHVKGSLNRJAAA 229

DB 168 DQLEHYIDNEKIQKLAFQTLVIGIDPKRGSLSYIIEMIEWMFGVHFVHFKGMYGMAQG 227

QY 230 MAQVIANGGEIHLNSETIESLIIENGAAGVKLOHGAELRG-DEVIINADFAHATHLVK 288

DB 228 LAQNKDLGVNIELNABIEQIIIDPKFKADAIKVGNDIRKEDKILCTADFPFSAESLMP 287

QY 289 P-GVLKYYTPENLKQREYSCSTFMYLGLD-KIYD-LPHHTVFAKDYTNIRNIFDNKT 345

DB 288 DFAPIKKYPHKLADLDYSCSAFLMTYIGIDIVTQVRLNLFVPSDDFGNIEELFEGE- 346

QY 346 LTDDFSFYVONASASDSDSLAPAGKSALYVLPMP--NNDSGLDW--QAHQCNVREQVLD 401

DB 347 LSYDPSIYVVVPAVADKSLAPAGKTIYVLMPTPELKTGSGIDWSEALTOQIKELIYRK 406

QY 402 LGARLGLDIRAHIECEKIITPQTWETDEHYKGFATSLSHKFSQMLYWRPNRFEELAN 461

DB 407 LATIEVFEDIKSHIVSETIFTDNDPEQTYHAKFGSAGLMPFLAQSNTYRPNQVSRDYK 466

QY 462 CYLVGGGTHPGSGLPTIYESARIS 485

DB 467 LYFAGASTHPGAGVPILTSAKIT 490

```

RESULT 13
D ADAL4542 standard; protein; 502 AA.
X ADAL4542;
T 06-NOV-2003 (first entry)
X Staphylococcus aureus CrtN protein SEQ ID NO:12.
X carotenoid biosynthetic enzyme; CBE;
X omega-aldehyde-functionalised carotenoid;
X omega-carboxy-functionalised carotenoid; C30 carotenoid;
X carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;
X CrtN.
X Staphylococcus aureus.
X WO2003068917-A2.
X 21-AUG-2003.
X 11-FEB-2003; 2003WO-US004150.
X 11-FEB-2002; 2002US-0355939P.
X (DUPO) DU PONT DE NEMOURS & CO E I.
X Cheng Q, Norton KC, Tao L;
X WPI; 2003-697524/56.
X N-PSDB; ADAL4541.
X New nucleic acid encoding carotenoid biosynthesis enzymes, useful for
X preparing functionalized carotenoids, e.g. pigments and antioxidants,
X also encoded polypeptides.
X Example 4; Page 108-111; 125pp; English.
X The present invention describes nucleic acid sequences encoding
X carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE
X polypeptides; (2) a chimeric gene (CG) containing a CBE linked to
X regulatory sequences; (3) host cell transformed with CG; (4) obtaining a
X nucleic acid that encodes a CBE; (5) products of method (4); (6)
X producing an omega-aldehyde-functionalised carotenoid (X); (7) producing
X an omega-carboxy-functionalised carotenoid (Y); (8) producing C30
X carotenoids (Z); and (9) regulating carotenoid biosynthesis in an
X organism. CBE sequences have antioxidant activity. CBEs can be used for
X microbial production of aldehyde- or carboxy-functionalised carotenoids
X and C30 carotenoids, and are potentially useful as pigments and
X antioxidants, e.g. as feed additives. The present sequence represents a
X CrtN CBE protein isolated from Staphylococcus aureus, which is used in
X the exemplification of the present invention.
X Sequence 502 AA;

```

```

Query Match 25.9%; Score 702.5; DB 6; Length 502;
Best Local Similarity 31.5%; Pred. No. 1.5e-61;
Matches 159; Conservative 108; Mismatches 196; Indels 41; Gaps 12;
QY 7 IIVGAGPGLCAGMLLSQKPKSVIFDKHAEIGGRNPINNGFTTGTFTLMMKVL 66
DB 3 IAVIGAGVTGLAAARIASQGEHTEIFKNNVNGKQKKGFTFDMGFTIWMMPDVY 62
QY 67 DEMFELCERSEDYLFPLSPMYRLLYDDRD-IFVSDRENMAELQRFDEGTDCYEQ 125
DB 63 KDVFYACKGKVEDYIELRQLRIYDYVDFPHDRDIIVPTDLAEQLQMLESTPGSTGFM 122
QY 126 FMOQEKRFNALYPCPTR-----DVSSLKSLFLSLLDLKALPFLAFPKSVNNL 173
DB 123 FLTDVYKVE-----IARYFLERTYRKPSDFYNMTSLVQAKLKL-----NHA 167

```

```

QY 174 GO----YENQEKMLAFECFQSKYLGKSPNECPALFTMLPVLEHEHYGYHVKGGLNRIARA 229
DB 168 DQLEHIDNEKIKLAFOTLITIGIDPKGPGSLISLIPIHMMFQVHFIKGMYGMAGC 227
QY 230 MAQVTAENGGEIHLNSESIIENGAAKGVKQHGAELEG-DEVIINADFAHAMTHLVK 288
DB 228 LAQMKDLGVNIELNABIEQIIIDPKFKRADAIVKNGDIKAFDKILCTADFFPSVAESLMP 287
QY 289 P-GVLKKTYPENLKOREVSCSTFMVLGLD-KIVD-LPHETIVFAKYDTTNIERNIDNKT 345
DB 288 DPAIKKYPKHKIADLDYSCSAFLMWIGIDIVTDQVRLENVIFSDDFRGNTEEIEGR- 346
QY 346 LTDFSFYVQNASDSDSLAPAKSALYVLVMP--NNDGGLDM--QAHCQNVREQLDT 401
DB 347 LSYDPSIYVYVPAVADKSLAPEKGTGIYVLMPPELKTGEGIDMSDEALTQOKETIYRK 406
QY 402 LGARLGLSDIRAHTECEKIITPOTWETDEHVYKATPSLCHKFSQMLYWRPNRPFELAN 461
DB 407 LATIEVPEDIKSHIVSETIFTFNDPFGTYHAKTSAPGLAPTLAQSNYYRPNQVSRDVKD 466
QY 462 CYLVGGGTHPGSLPTIYESARIS 485
DB 467 LYFAGASTHPCAGVPIVLSAKIT 490

```

RESULT 14

```

AAR07466
ID AAR07466 standard; protein; 492 AA.

```

```

XX AC AAR07466;
XX AC
XX DT 24-OCT-2003 (revised)
XX DT 28-JAN-1991 (first entry)
XX DE Polypeptide with enzymatic activity for the conversion of phytoene into
XX DE lycopene.
XX KW Carotenoid biosynthesis; vitamin A; cancer; food coloring.
XX OS Pantoea ananatis.
XX PN EP393690-A.
XX PD 24-OCT-1990.
XX PF 20-APR-1990; 90EP-00107493.
XX PR 21-APR-1989; 89JP-00103078.
XX PR 05-MAR-1990; 90JP-00053255.
XX PA (KIRI) KIRIN BEER KK.
XX PI Misawa N, Kobayashi K, Nakamura K;
XX DR WPI; 1990-322212/43.
XX DR N-PSDB; AAQ06296.

```

```

PT DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn.
PT of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.
XX Claim 4; Fig 4; 40pp; English.
XX Gene products are useful for the synthesis of carotenoids, useful as food
XX coloring, vitamin A precursor, and possibly in prevention of cancer. See
XX also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)
XX Sequence 492 AA;

```

```

Query Match 25.8%; Score 700.5; DB 2; Length 492;
Best Local Similarity 31.5%; Pred. No. 2.9e-61;
Matches 157; Conservative 103; Mismatches 213; Indels 25; Gaps 10;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 24.4213 Seconds
(without alignments)
3837.172 Million cell updates/sec

file: US-09-941-947A-20

Effect score: 1510
sequence: 1 MSKUKAYLTVCQVERALD.....FGSEADLLRLSYLIERTH 237

scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	831	55.0	295	16 Q9HWY4	Q9HWY4 pseudomonas
2	811	53.7	295	16 Q8BQC6	Q8BQC6 pseudomonas
3	800.5	53.0	300	2 Q9S5F0	Q9S5F0 synchococ
4	791	52.4	235	16 Q8B9Q0	Q8B9Q0 pseudomonas
5	787.5	52.2	237	16 Q8ZVD4	Q8ZVD4 nitrosomona
6	729	48.3	298	16 Q9USM0	Q9USM0 neisseria m
7	715	47.4	231	16 Q8P7L1	Q8P7L1 xanthomonas
8	707	46.8	298	16 Q8DFA4	Q8DFA4 vibrio vuln
9	706	46.8	291	16 Q8PIY3	Q8PIY3 xanthomonas
10	706	46.8	294	16 Q8RT9	Q8RT9 vibrio para
11	705.5	46.7	298	16 Q8ZRD0	Q8ZRD0 salmonella
12	705.5	46.7	299	16 Q8ZEX2	Q8ZEX2 salmonella
13	704	46.6	294	16 Q8XTE2	Q8XTE2 vibrio chol
14	702.5	46.5	306	16 Q8ZC46	Q8ZC46 yersinia pe
15	695.5	46.1	299	16 Q8XE75	Q8XE75 escherichia
16	695.5	46.1	308	16 Q7VLUL	Q7VLUL haemophilus

ALIGNMENTS

RESULT 1

Q9HWY4 Q9HWY4 PRELIMINARY; PRT; 295 AA.
ID Q9HWY4
AC Q9HWY4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Geranyltransferase.
GN IS OR PA0403.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
CX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltz L., Tolentino R., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964 (2000).
DR EMBL; AB004821; AAC07430.1; --
DR PIR; F83139; F83139.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00348; polyprenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
DR Transferase; Complete proteome.
SQ SEQUENCE 295 AA; 31464 MW; EFDFFAD66F5C126 CRC64;

Query Match

55.0%; Score 831; DB 16; Length 295;


```

01-JUN-2003 (TREMELrel. 24, Created)
01-JUN-2003 (TREMELrel. 24, Last sequence update)
01-OCT-2003 (TREMELrel. 25, Last annotation update)
Geranyltransferase.
ISPA OR PSPT06699.
Pseudomonas syringae (pv. tomato).
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=323;
[1]
SEQUENCE FROM N.A.
STRAIN=DC3000;
Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
Berry K., Utterback T., Van Aken S., Feildlyum T., Gwin M.,
Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
White O., Fraser C., Collmer A.,
"Complete sequence of Pseudomonas syringae.";
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
EMBL; AE016558; AA054241.1; -
TIGR; PSPT06699; -
GO; GO:0016740; F:transferase activity; IEA.
GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
InterPro; IPR000092; Polyprenyl synth.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synth; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
Transferase; Complete proteome.
Q SEQUENCE 295 AA; 31255 MW; 86DA1E6092FB501A CRC64;

Query Match 52.4%; Score 791; DB 16; Length 295;
Best Local Similarity 56.8%; Pred. No. 2e-55;
Matches 167; Conservative 36; Mismatches 91; Indels 0; Gaps 0;

Y 4 LKAYLTVCQERVALDARLPANILPQTLHQAMRYSVLNGGKRTPLTYATGALGLP 63
b 2 IASVQASQTRVNAALGEGFQAPSPSLRYEMRYSVWNGKVRPLLAYAACALGV 51
Y 64 ENVLDPACAVEIHVYSLIHDDLPAMDNDLRRGKPTCHKYDEATAILAGDALQALAF 123
b 62 AEDANGAACAVELIHAYSLVHDDLPAMDNDLRRGQPTTHKAFDEACATLAGDGLQSLAF 121
Y 124 EVLADPGITVDAPARLKMILTALTRASSQGMVGGQAIIDLSGVGRKLTLPENLHHIHT 183
b 122 TALLAPHLSSRNATRLQWSTLAAAGAGMVGQAIIDLSGVGLKLDQALLEYMHRHT 181
Y 184 GALIRASVNLAAALSKPDLTCTVAKKLDHYAKCIGLSFQVKDILDIADTATLGTQKGD 243
b 182 GALIEASVRLGALASQADQARLDALQVYARAVGLAFQVDDILDVSDTATLGTQKQAD 241
Y 244 INDKPTYPALLGMCAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297
b 242 IARDKPTYPALLGLDAKGYALELRDQALALSLFDTTAEPLRELYIVERRH 295

RESULT 5
82VD4 PRELIMINARY; PRT; 297 AA.
C Q82VD4
T 01-JUN-2003 (TREMELrel. 24, Created)
T 01-JUN-2003 (TREMELrel. 24, Last sequence update)
T 01-OCT-2003 (TREMELrel. 25, Last annotation update)
Polyprenyl synthetase (EC 2.5.1.10).
ISPA OR NE1160.
Pseudomonas europaea.
Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
Nitrosomonadaceae; Nitrosomonas.
NCBI_TaxID=915;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 19718 / IFO 14298;
MEDLINE=22586410; PubMed=12700255;

```

```

RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.C., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.,
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773 (2003).
DR EMBL; BX321860; CAD85071.1; -
DR GO; GO:0004337; F:geranyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 297 AA; 32448 MW; A22BC33365F3699 CRC64;

Query Match 52.2%; Score 787.5; DB 16; Length 297;
Best Local Similarity 56.9%; Pred. No. 3.9e-55;
Matches 160; Conservative 40; Mismatches 80; Indels 1; Gaps 1;

QY 15 VERALDARLPANILPQTLHQAMRYSVLNGGKRTPLTYATGALGLPENVLDPACAV 74
DB 16 VETCLEXHLPEITNCAPARLDHMYRYVLGGKVRPLLSFAGELSGADKTHATIAAAV 75
QY 75 EFHIVYSLIHDDLPAMDNDLRRGKPTCHKYDEATAILAGDALQALAFVLANDPQITV 134
DB 76 ELIHVYSLVHDDLPAMDNDLRRGKPTCHKYDEPTALLWDSLSQSLAFOLL-TETNLTE 134
QY 135 DAPARLKMILTALTRASSQGMVGGQAIIDLSGVGRKLTLPENLHHIHTKALIRASVNL 194
DB 135 DPHVQLEWVRHLAFAGSRGMAGGQAIIDLSVGRKLTSLPELFMHHIHTKALIRAAVILG 194
QY 195 ALSKPDLDTCVAKKLDHYAKCIGLSFQVKDILDIADTATLGTQKGDINDKPTYPAL 254
DB 195 ARCGRNLDLTCVAKKLDHYAKCIGLSFQVKDILDIADTATLGTQKGDINDKPTYPAL 254
QY 255 LGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIER 295
DB 255 LGTKQARELAHELQREATGVINQFGSEALRLQQTDFIVQR 295

RESULT 6
Q9JSMO PRELIMINARY; PRT; 298 AA.
ID Q9JSMO
AC Q9JSMO;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Geranyltransferase (EC 2.5.1.10).
GN ISPA OR NMA2226.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=22491; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagsis K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506 (2000).
DR EMBL; AL162758; CAB85437.1; -
DR PIR; E81796; E81796.
DR GO; GO:0004337; F:geranyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

```


129 DPGITVDAPARLKMITALTRASGQGMVGGQALDGLSVGRKLTLPLENHGHKTKALIR 188
130 GTLSADGETQVAMQALQASGAQGCICGOALDLAENRAVLEELERHKTGALIR 189
189 ASVNLALSKPDLTVCVAKLDHYAKIGLSFOVKDDILDIEDATTLGKTQCKDINDK 248
190 CAVRLGALAGEKGLALPOLKEYSTAGLAFQVDDILDIDSTTLGKPGQSDQELNK 249
249 PTPALGMAGAKOKAQELHEQAVESITGFGSEADLLRELSLYIER 295
250 STYPSLLGLEGAKETLHLALQALEATPYNTQHLERFARVWER 296

RESULT 9

Q87RT9 PRELIMINARY; PRT; 291 AA.
AC Q87RT9, 24, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE Geranyltransferase.
GN VP0687.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Vibrionaceae; Xanthomonas.
CX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferris J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Camuara A.F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Lemos E.G.M.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Martins-Rossi N.H.M.,
RA Locali E.C., Machado M.A., Madeira A.B.N., Martinez-Rossi N.H.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
T "Comparison of the genomes of two Xanthomonas pathogens with differing
T host specificities."
T Nature 417:459-463(2002).

EMBL: AB013916; AM37607.1; --
R GO: GO:0016740; F:transferase activity; IEA.
R GO: GO:0008299; P:isoprenoid biosynthesis; IEA.
R InterPro: IPR000092; Polyprenyl synth.
R InterPro: IPR008949; Terpenoid synth.
R Pfam: PF00348; polyprenyl synth_1
R PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
R PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
R Transferase; Complete proteome.
Q SEQUENCE 291 AA; 30516 MW; 404385EDFE8BF67 CRC64;

Query Match 46.8%; Score 706; DB 16; Length 291;
Best Local Similarity 51.7%; Pred. No. 1.4e-48;
Matches 148; Conservative 44; Mismatches 86; Indels 8; Gaps 3;

14 RVERDALPRLPAENILPQTLHQAMRYSVLNGGKTRPLLTATGQALGLPENVL 73
12 RTERSLEAGLPSATHAPORLHAARHVALVGGKRWPLLVASGALFGAEDQLTPAVA 71
74 VFHFVYSLIHDDLPAMDNDLRRCKPTCHKAYDEATLADLQALAFVNLDPGIT 133
72 VELLFAYSLVHDDLPAMDNDLRRCKPTCHKAYDEATLADLQALAFVNLDPGIT 128
134 VDAPARLKMITALTRASGQGMVGGQALDGLSVGRKLTLPLENHGHKTKALIRASVNL 193

129 ASAEIRVGMQSLATAAGAGMCGQALDIDATQLOSLQHLQRMHAKTKGALIRAVRM 188
194 AALSKPDLTVCVAKLDHYAKIGLSFOVKDDILDIEDATTLGKTQCKDINDKPT 251
189 GALTG--GAALADQORLDFADALGAFQVDDILDIDSTTLGKPGQSDQELNK 245
252 PALLGMAGAKOKAQELHEQAVESITGFGSEADLLRELSLYIER 297
246 PALLGMAGAKAKLAELAAARHMDVLPQVGPGETLATLARFAVNR 291

RESULT 10

Q87RT9 PRELIMINARY; PRT; 294 AA.
AC Q87RT9, 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE Geranyltransferase.
GN VP0687.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
CX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIND 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nishijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749(2003).
DR EMBL: AP005075; BAC8950.1; --
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro: IPR000092; Polyprenyl synth.
DR InterPro: IPR008949; Terpenoid synth.
DR Pfam: PF00348; polyprenyl synth_1
DR PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
DR Transferase; Complete proteome.
SQ SEQUENCE 294 AA; 32337 MW; A1C1924A66774FD0 CRC64;

Query Match 46.8%; Score 706; DB 16; Length 294;
Best Local Similarity 52.1%; Pred. No. 1.4e-48;
Matches 150; Conservative 41; Mismatches 97; Indels 0; Gaps 0;

QY 8 LTVQBRVERDALPRLPAENILPQTLHQAMRYSVLNGGKTRPLLTATGQALGLPENVL 67
DB 5 LTVSFRNNQNLNMLEQLPYQEQPLIQAMKYGLLGGKRWPLLVITGQMLGCKPEDL 64
QY 68 DAPACAVEFHVYSLIHDDLPAMDNDLRRCKPTCHKAYDEATLADLQALAFVNL 127
DB 65 DTPASAEICIHAYSLIHDDLPAMDNDLRRCKPTCHKAYDEATLADLQALAFVNL 124
QY 128 NDPGITVDAPARLKMITALTRASGQGMVGGQALDGLSVGRKLTLPLENHGHKTKALI 187
DB 125 DGPLSPAENQVRNMLKALHSSGANGCMVCGALDGLAENQVSLAEHIEHKTGALI 184
QY 188 RASVNLALSKPDLTVCVAKLDHYAKIGLSFOVKDDILDIEDATTLGKTQCKDIND 247
DB 185 DCVAVKLGAALAGEKGLVPLHRYSKAIGLAFQVDDILDIDSTTLGKPGQSDQELNK 244
QY 248 KPTYPALLGMAGAKOKAQELHEQAVESITGFGSEADLLRELSLYIER 295
DB 245 KSTYPSLLGLEGAKETLHLQALEATPYNTQHLERFARVWER 292

RESULT 11

Q8ZED0

```
ID Q82RDO PRELIMINARY; PRT; 299 AA.
AC Q82RDO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Geranyltransferase (farnesylidiphosphate synthase) (EC
DE 2.5.1.10)
GN ISPA OR SFW0423.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]_TaxID=602;
RN SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856 (2001).
DR EMBL; AE008715; AAL19377.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyrenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00348; Polyrenyl_synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 299 AA; 3196 MW; 5B9BF79CCCF0FC3C CRC64;

Query Match 46.7%; Score 705.5; DB 16; Length 299;
Best Local Similarity 52.7%; Pred. No. 1.6e-48;
Matches 156; Conservative 39; Mismatches 94; Indels 7; Gaps 3;

QY 3 KKKAVLTVCQRRVERALDARLPAENILPQTLHQAMRYSLVNGGKRTPLLTATGQALGL 62
DB 6 QIQACVTOANQALSRFI-APLPFQ--TPVEAMQYCALLGGKRLPFLVYATGQMGV 61
QY 63 PENVLDPACAVEFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALA 122
DB 62 STATLDAPAAAVECHAYSLIHDLPAMDNDLRRGLPTCHKIPEANAILAGDALQTLA 121
QY 123 FEVLANDPGITVDAPARLKMITALTRASSGQGMVGGQALDLSVGRKLTPELENMHIHK 182
DB 122 FAIISDAPMPVADRDRIAMIAELANASGIAGMCGQALDLAEGQRTILDALRIHRHK 181
QY 183 TGAIRASVNLAAALSKPDLDTCAVKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQK 242
DB 182 TGAIRAAVRLGALSAGDKGRNTLPILDRVAESIGLAFQVQDDILDVVGDTATLGKQGA 241
QY 243 DIDNDKPTYPALLGKAGAKQAKQELHEQAVESLTFGSE--DILRELSLYIER 295
DB 242 DQQLGKSTYPALLGLEQARNKARDLIEDARQSLHQAQSLDTSALEALANYIIOR 297

RESULT 12
Q82RDX2 PRELIMINARY; PRT; 299 AA.
ID Q82RDX2
AC Q82RDX2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Geranyltransferase.
DE STY0462 OR ISPA OR T2440.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
```

```

RN SEQUENCE FROM N.A.
RP STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852 (2001).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.P., Rose D.J.,
RA Burland V., Kodyourami V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337 (2003).
DR EMBL; AL627266; CAD08879.1; -.
DR EMBL; AE016842; AA070030.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyrenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00348; Polyrenyl_synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 299 AA; 32016 MW; 512BF82369EA568C CRC64;

Query Match 46.7%; Score 705.5; DB 16; Length 299;
Best Local Similarity 52.7%; Pred. No. 1.6e-48;
Matches 156; Conservative 39; Mismatches 94; Indels 7; Gaps 3;

QY 3 KKKAVLTVCQRRVERALDARLPAENILPQTLHQAMRYSLVNGGKRTPLLTATGQALGL 62
DB 6 QIQACVTOANQALSRFI-APLPFQ--TPVEAMQYCALLGGKRLPFLVYATGQMGV 61
QY 63 PENVLDPACAVEFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALA 122
DB 62 STATLDAPAAAVECHAYSLIHDLPAMDNDLRRGLPTCHKIPEANAILAGDALQTLA 121
QY 123 FEVLANDPGITVDAPARLKMITALTRASSGQGMVGGQALDLSVGRKLTPELENMHIHK 182
DB 122 FAIISDAPMPVADRDRIAMIAELANASGIAGMCGQALDLAEGQRTILDALRIHRHK 181
QY 183 TGAIRASVNLAAALSKPDLDTCAVKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQK 242
DB 182 TGAIRAAVRLGALSAGDKGRNTLPILDRVAESIGLAFQVQDDILDVVGDTATLGKQGA 241
QY 243 DIDNDKPTYPALLGKAGAKQAKQELHEQAVESLTFGSE--DILRELSLYIER 295
DB 242 DQQLGKSTYPALLGLEQARNKARDLIEDARQSLHQAQSLDTSALEALANYIIOR 297

RESULT 13
Q8KTL2 PRELIMINARY; PRT; 294 AA.
ID Q8KTL2
AC Q8KTL2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Geranyltransferase.
DE VC0890.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
```

```

X NCBI_TaxID=666;
P [1]
N SEQUENCE FROM N.A.
C STRAIN=BI Tor N16961 / Serotype O1;
C MEDLINE=20406833; PubMed=10952301;
X Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
A Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
A Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
A Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
A McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
A Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
A Fraser C.M.;
T "DNA sequence of both chromosomes of the cholera pathogen Vibrio
L cholerae.";
L Nature 406:477-483(2000).
L EMBL; AE004173; AAF94052.1; -.
R PIR; A82267; A82267.
R TIGR; VCC890; -.
R GO; GO:0016740; F:transferase activity; IEA.
R GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
R InterPro; IPR008092; Polyprenyl synth.
R Pfam; PF00348; polyprenyl synth.
R PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
R PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
W Transferase; Complete proteome.
Q SEQUENCE 294 AA; 32259 MW; E7D33C0FD935CB96 CRC64;

Query Match 46.6%; Score 704; DB 16; Length 294;
Best Local Similarity 51.2%; Pred. No. 2e-48;
Matches 149; Conservative 45; Mismatches 91; Indels 6; Gaps 2;

/ 8 LTVQQRVERALDA---RLPAENILPQTILQAMRYSVLNGKRTPLTYATGQALGPE 64
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 5 LSSYQQRNNQQLDMLNRPFTQL---PLIEAMRYGLLGGKRRAPYLYVTGQMLGCEL 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 65 NVLDAPACAVEFHVYSLIHDDLPAWDDNLLRGGKPTCHKAYDEATAILAGDALQALFE 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 62 SLDLTPASAVECHAYSLIHDDLPAWDDNLLRGGKPTCHKAYDEATAILAGDALQALFE 121
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 125 VLANDPGITVDAPARKMITALTATRASGQGVGGQAIIDLSGVGRKLTPELENMHIHRTG 184
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 122 ILAEGDLSAAGETQRYAMQLAAEASAGQCMGLGQALDLAAENRILSLELETHRNKNG 181
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 185 ALIRASVNLAAASKPDLPTCVAKLDHYKACIGLSFQVKKDDILDIEDATATGKTGKDI 244
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 182 ALMRCAIRLGLAAEKGKRAMPLHLDRYAEAVGLAFQVQDDILOIISTETLGPQGSQDQ 241
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 245 DNDKPTYPALLGAGAKAKQAQELHQAVESLTGFGSEADLLRELSLYIERTH 295
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 242 ELNKSTYPALLGLEGAQQAHTLLQBALLALEAIPYNTHELEEFARYVVER 292
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Kariyshev A.V.,
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds W., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RC MEDLINE=22137863; PubMed=12142430;
RX Deng W., Burland V., Plunkett G., Boutin A., Mayhew G.F., Liss P.,
RX Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AJ414155; CAC92411.1; -.
DR EMBL; AE013704; AAM84590.1; -.
DR PIR; AH0385; AH0385.
DR GO; GO:0004337; F:geranyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 306 AA; 32919 MW; 84F61276DFB6B106 CRC64;

Query Match 46.5%; Score 702.5; DB 16; Length 306;
Best Local Similarity 52.5%; Pred. No. 2.8e-48;
Matches 157; Conservative 38; Mismatches 95; Indels 9; Gaps 3;

QY 5 KAYLTVCQQRVERAL---DARLPAENILPQTILQAMRYSVLNGKRTPLTYATGQALG 61
DB 11 KQOLAARQVRVQALDLFTAPLPGN---SNLEAMRYGAVIGGKRLRPLYVYATGQMGF 67
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 LPENVLDAPACAVEFHVYSLIHDDLPAWDDNLLRGGKPTCHKAYDEATAILAGDALQAL 121
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 LSLANDAPAAAEICHHAYSLIHDDLPAWDDNLLRGGKPTCHKAYDEATAILAGDALQAL 127
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 APEVLANDPGITVDAPARKMITALTATRASGQGVGGQAIIDLSGVGRKLTPELENMHIH 181
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 128 AFSILAEAMPVADVADKRLSMIIELAAQASGAGMAGQALDLAEATRPVLLGDLRQIHRH 187
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 KTGALIRASVNLAAASKPDLPTCVAKLDHYKACIGLSFQVKKDDILDIEDATATGKTGQ 241
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 KTGALIRAAVRLGAQAAGEAGKRTLLSLDRYAEATGLAFQVQDDILDVIGTATIGKRG 247
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 242 KDIDNDKPTYPALLGAGAKAKQAQELHQAVESLTGFGSEA---DLLRELSLYIERTH 297
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 SDQQLGKSTYPALLGUDCAQTAMDLYQELAEALDELAKQSTNTSLQALAFIIRNN 306
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

Q8XE75 PRELIMINARY; PRT; 299 AA.
AC Q8XE75;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Geranyltransferase (farnesyl diphosphate synthase).
GN ISPA OR YP03176 OR Y1009.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;

```

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RA Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
XX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RA DNA Res. 8:11-22(2001).
DR EMBL; AE005221; AAC54771.1; -;
DR EMBL; AF002551; BAB33896.1; -;
DR FIR; C90688; C90688.
DR FIR; G85538; G85538.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyrenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; Polyrenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
DR Transferase; Complete proteome.
CW SEQUENCE 299 AA; 32131 MW; 15BC6BECBE7FB0CA CRC64;
SQ
Query Match 46.1%; Score 695.5; DB 16; Length 299;
Best Local Similarity 52.4%; Pred. No. 9.9e-48;
Matches 154; Conservative 35; Mismatches 96; Indels 9; Gaps 3;
2Y 8 LTVQCHVERALD---ARLPAENILPOTLHOAMRYSVLNGCKRTPRLTYATQALGLPE 64
DB 7 LEACVKQANQALSRIAPLPQN---TPVETWQYGALLGKRLRPFLVYATGHMFGVST 63
2Y 65 NVLDAPACAVEFTHVYSLIHDLDPAMDNDLIRGKPTCHKAYDEATAILAGDALQALAFE 124
DB 64 NTLDPAAAAVECIHAYSLIHDLDPAMDNDLIRGKPTCHKAYDEATAILAGDALQALAFE 123
2Y 125 VLANDPGITVDAPARLKMITALTASGSGQWVGQALDEASGVGRKLTLPENWHIKTG 184
DB 124 ILSDDMPFVSDRDRIISMISLASGSIAGMCGGQALDLDAGKHVFLDALRIHRKKG 183
2Y 185 ALIRASVNLALSKPDLPTCVAKLDHYAKICIGLSFQVKDDILDIEADTATLTKTQKDI 244
DB 184 ALIRAAVRLGALSAGDKGRALPVLIDKYAESIGLAFQVQDDILDVWGDATLTKRQGAQ 243
2Y 245 DNDKETYALLGMCAKQAKQELHEQAVESLTGPGSEA---DLRLSLYIIER 295
DB 244 QLGRSTYFALLGLEQARKKAQDLIDARQSLKQLAEQSIDTSALEALADYIIQR 297

Search completed: February 29, 2004, 14:50:36

Job time : 31.4213 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 8.26685 Seconds

(without alignments)
3455.835 Million cell updates/sec

US-09-941-947A-20

title:

effect score: 1510

sequence: 1 MSKLKAVLTVCQVERALD.....FGSEADLLRELSLYIERTH 297

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

real number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Maximum Match 0%

Listing first 45 summaries

Database: PIR 78:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	831	55.0	295	2 F83139	Geranyltransferase
2	729	48.3	298	2 E81796	Geranyltransferase
3	705.5	46.7	299	2 AG0554	Geranyltransferase
4	704	46.6	294	2 A82267	Geranyltransferase
5	702.5	46.5	306	2 AH0385	Geranyltransferase
6	695.5	46.1	299	2 C90688	Geranyltransferase
7	695.5	46.1	299	2 G85538	Geranyltransferase
8	691.5	45.8	299	2 JQ0665	Geranyltransferase
9	677	44.8	295	2 C64123	Geranyltransferase
10	675	44.7	291	2 D82778	Geranyltransferase
11	668	44.2	259	2 F81217	Geranyltransferase
12	606.5	40.2	302	2 S74538	Geranyltransferase
13	604	40.0	371	2 F84334	Geranyltransferase
14	596	39.5	369	2 S53722	Geranyltransferase
15	590.5	39.1	366	2 T10452	Geranyltransferase
16	586	38.8	309	2 AE1833	Geranyltransferase
17	583	38.6	332	2 T10875	Geranyltransferase
18	578	38.3	332	2 I40213	Geranyltransferase
19	577	38.2	297	2 JX0257	Geranyltransferase
20	575	38.1	357	2 T09966	Geranyltransferase
21	567.5	37.6	347	2 E84566	Geranyltransferase
22	565	37.4	294	2 E83997	Geranyltransferase
23	565	37.4	372	2 G84566	Geranyltransferase
24	557.5	36.9	304	2 A13285	Geranyltransferase
25	552.5	36.6	316	2 T11021	Geranyltransferase
26	541.5	35.9	294	2 D87505	Geranyltransferase
27	537	35.6	293	2 AC1245	Geranyltransferase
28	531	35.2	272	2 A89961	Geranyltransferase
29	531	35.2	376	2 T02429	Geranyltransferase

RESULT 2

E81796

Geranyltransferase (EC 2.5.1.10) NMA2226 [imported] - Neisseria meningitidis (strain E81796)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: E81796

30 529 35.0 293 2 AG1607 geranyltransferase
31 522.5 34.6 300 2 T06969 farnesyltransferase
32 522 34.6 378 2 S71230 geranyltransferase
33 517.5 34.3 293 2 A89932 hypothetical protein
34 517 34.2 304 2 F97685 geranyltransferase
35 517 34.2 335 2 AH2910 geranyltransferase
36 501 33.2 282 2 B84984 geranyltransferase
37 488.5 32.4 285 2 A85732 geranyltransferase
38 476.5 31.6 289 2 D97156 probable geranyltransferase
39 446.5 29.6 291 2 G95139 geranyltransferase
40 445.5 29.5 291 2 F98007 geranyltransferase
41 440.5 29.2 222 2 S71231 geranyltransferase
42 399 26.4 289 2 S04407 phytoene synthase
43 391 25.9 303 2 A64636 geranyltransferase
44 379 25.1 281 2 B81261 geranyltransferase
45 375 24.8 303 2 G71878 geranyltransferase

ALIGNMENTS

RESULT 1

F83139

Geranyltransferase PA043 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: F83139

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bz
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: F83139

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-295 <STO>

A;Cross-references: GB:AE004821; GB:AE004091; NID:g9950236; PIDN:AAG07430.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: is; PA0403

Query Match 55.0%; Score 831; DB 2; Length 295;

Best Local Similarity 60.3%; Pred. No. 2.6e-58;

Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;

Qy 4 LKAYLTVCCQVERALDARLPANILPOTLHOAMRYSYLNGGKTRPLLTATGQALGLP 63

Db 2 IAAQACQARVDALDALFVAPRELQELYENARYSYWNGGKRVRELLAYAAACEALGA 61

Qy 64 ENVLDPACAVFEHIVYSLIHDDLPAMDNDLRRKPTCHAYDEATAILAGDALQALAP 123

Db 62 PORADAAACAVELIHAYSLVHDDLPAMDNDLRRKQPTTHAFDEATAILAADGLQALAF 121

Qy 124 EVLANDPGITVPAPARLKMITALTRASSQGVGGQALDLSGVGRKLTLPLENNHIHKT 183

Db 122 EVLADTRRNQPHQVAVCLEMLTFLARAGSAGVGGQALDLSGVVALQQAILEVWHRKT 181

Qy 184 GALIRASVNLAAALSKPDLDTVCVAKKLHRYAKTIGHSFQVXDDILDIEDATATLGTQCKD 243

Db 182 GALIEASVRLGALAGRAEPASIAALRYEAYEAIGLAFQVQDDILDVSDTATLGTQCKD 241

Qy 244 INDNKTPTPALIGMAGAKQKQELHQAVESLTGFGSEADLLRELSLYIER 295

Db 242 QAHNKPTYPALIGLEAAKGYALELRDLAALDGPFPSPADPLRQLARYIVER 293

R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
Holtroyd, S.; Jagsis, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: AB1775; MUID:2022556; PMID:10761919
A:Accession: E81796
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <PAR>
A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85437.1; PID:g738064
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
C:Gene: *ispA*; NMA2226
C:Superfamily: dimethylallyltransferase
C:Keywords: transferase

Query Match 48.3%; Score 729; DB 2; Length 298;
Best Local Similarity 51.4%; Pred. No. 3.1e-50;
Matches 151; Conservative 44; Mismatches 97; Indels 2; Gaps 1;

QY 4 LKAYLTVCQVEREALDARLPANILPQTLHQAMYSVLNGKRTPLLTATGQALGLP 63
DB 7 LKAOQQAQATOLLERFLPSENEIPTLHEARYAALDGGKRLPMLVLAASELGEAV 66
QY 64 ENVLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123
DB 67 HEAVQMAAIEIMIHAYSLVHDDPAMDNDLRRGKPTCHKVYGEATALLTGDAQTQAF 126
QY 124 EVLANDPGITVDAPARLKMITALTRASQGMVGGQALDLSVGRKLTPELENMHIKT 183
DB 127 DVLSRP--TELPAAQALAMLSVLAKAGSGMAGQALDLANQKWAQTDLEFQMSLKT 184
QY 184 GALIRASVNLAAALSKPDLDTCAKLDHVAKIGLSFQVKKDILDEADTATLGTQKDI 244
DB 185 GALIRASVNLAAALSKPDLDTCAKLDHVAKIGLSFQVKKDILDEADTATLGTQKDI 244
QY 244 INDNDKPTYPALLGMAGAKQKQAEHQAQVSLTGFGEADLLRELSLYIIR 297
DB 245 ADNDKPTYPALLGMAGAKQKQAEHQAQVSLTGFGEADLLRELSLYIIR 297

RESULT 3
AG0554
geranyltransferase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: This species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0554
R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
Ch., T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
Ch., T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08879.1; PID:g16501691; GSPDB:GN00176
C:Genetics:
C:Gene: STY0462
C:Superfamily: dimethylallyltransferase

Query Match 46.7%; Score 705.5; DB 2; Length 299;
Best Local Similarity 52.7%; Pred. No. 2.2e-48;
Matches 156; Conservative 39; Mismatches 94; Indels 7; Gaps 3;

QY 3 KLAYLTVCQVEREALDARLPANILPQTLHQAMYSVLNGKRTPLLTATGQALGL 62
DB 6 QLOACVTQANALSEPI-APLPFQ--TPVVEAMQVALLGGRLRPPLVYATGQVFGV 61
QY 63 PENVLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALA 122

DB 62 STATLDAPAAVECHAYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALA 121
QY 123 FEVLANDPGITVDAPARLKMITALTRASQGMVGGQALDLSVGRKLTPELENMHIK 182
DB 122 FTIISDAPWPEVADRDRITAMIAELANASGIAGMGGQALDLAARGQRTITDLAERIHK 181
QY 183 TGALIRASVNLAAALSKPDLDTCAKLDHVAKIGLSFQVKKDILDEADTATLGTQK 242
DB 182 TGALIRASVNLAAALSKPDLDTCAKLDHVAKIGLSFQVKKDILDEADTATLGTQK 241
QY 243 DIDNDKPTYPALLGMAGAKQKQAEHQAQVSLTGFGEADLLRELSLYIIR 295
DB 242 DQQLGKSTYPALLGLEQAKRKARDLIEDARQSLHQAQAQSLDLSALEALANYIIR 297

RESULT 4
AB2267
geranyltransferase VC0890 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: AB2035; MUID:20406833; PMID:10952301
A:Accession: AB2267
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <HEI>
A:Cross-references: GB:AE004173; GB:AE003852; NID:g9655341; PIDN:AAF94052.1; GSPDB:GN00
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0890
A:Map position: 1
C:Superfamily: dimethylallyltransferase

Query Match 46.6%; Score 704; DB 2; Length 294;
Best Local Similarity 51.2%; Pred. No. 2.9e-48;
Matches 149; Conservative 45; Mismatches 91; Indels 6; Gaps 2;

QY 8 LTVCOEVERALDA---RLPAENILPQTLHQAMYSVLNGKRTPLLTATGQALGLPE 64
DB 5 LSSVQQRNNQQLDQWLNRPPTQL---PLIEAMRYGULLGKGRAPYLVITGQMLCEL 61
QY 65 NYLDAPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 124
DB 62 SDLDTPASAVCHAYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 121
QY 125 VLANDPGITVDAPARLKMITALTRASQGMVGGQALDLSVGRKLTPELENMHIKTG 184
DB 122 ILAEGDLISAAGETORVAMLOALAEASGACQKCLGQALDLAENLISLELETIHRNKTG 181
QY 185 ALIRASVNLAAALSKPDLDTCAKLDHVAKIGLSFQVKKDILDEADTATLGTQKDI 244
DB 182 ALMRCAIRLGAALAGEKGRAMPHLDRYAEAVGLAFQVQDDILDIISDTETLGRQSSDQ 241
QY 245 DNDKPTYPALLGMAGAKQKQAEHQAQVSLTGFGEADLLRELSLYIIR 295
DB 242 ELNKSTYPALLGLEQAKRKARDLIEDARQSLHQAQAQSLDLSALEALANYIIR 292

RESULT 5
AB0385
geranyltransferase (EC 2.5.1.10) [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AB0385
R.Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-farrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,


```
ature 413, 523-527, 2001
;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
;Reference number: AB0001; MUID:21470413; PMID:11586360
;Accession: AH0385
;Status: preliminary
;Molecule type: DNA
;Residues: 1-306 <KUR>
;Cross-references: GB:AL590842; PIDN:CAC92411.1; PID:gl5981114; GSPDB:GN00175
;Genetics:
;Gene: ispa
;Superfamily: dimethylallyltranstransferase
;Keywords: transferase

Query Match 46.1%; Score 702.5; DB 2; Length 306;
Best Local Similarity 52.5%; Pred. No. 1.4e-48;
Matches 157; Conservative 38; Mismatches 95; Indels 9; Gaps 3;

y 5 KAVLTWCQRRVERAL---DARLPAENILPQTLHQAMRYSVLNGGKTRPLLTATGQALG 61
b 11 KQGLAAHQQRVQNALDFFAPLPGN---SNLEAMRYGAVTGGKRLPVLVYATGQWFG 67
y 62 LPENVLDAPACAVEFHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQAL 121
b 68 LSLANLDAFAAECHAYSLIHDDLPAMDNDLRRGKPTCHKVGEAHAILAGDALQTL 127
y 122 AFEVLANDPGITVDAPARLKMITALTRASSGQMGVGGQIDLGSGVRKLTLPLEENMH 181
b 128 AFSILAEAPWPDYADVADKRLSMITELAQASGAACMCAGQALDLEASTRPVLLGDLEQIHRH 187
y 182 KTCALIRASVNLAAALSKPDLDTCAKLDHYAKCIGLSFQVNDLIDLEADTATLGTQ 241
b 188 KTCALIRAVRLGAQAGAGRKTLISLDRYAEAGLAFQVQDDILDVGTATLGTQ 247
y 242 KQINDKPTYPALLGWAGAKQAQELHEQAVESLTGFGSEA---DLLRELSLYIERTH 297
b 248 SDQQLGKSTYPALLGLDCAQTRAWDLYQEALEALQAKSYNTDSQALAFRIERN 306

RESULT 6
90688
eranyltranstransferase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05
;Species: Escherichia coli
;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
;Accession: C90688
;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
asawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
NA Res. 8, 11-22, 2001
;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
;Reference number: A95629; MUID:21156231; PMID:11258796
;Accession: C90688
;Status: preliminary
;Molecule type: DNA
;Residues: 1-299 <HAY>
;Cross-references: GB:BA000007; PIDN:BA033898.1; PID:gl3359932; GSPDB:GN00154
;Experimental source: strain O157:H7, substrain RIMD 0509952
;Genetics:
;Gene: ECs0475
;Superfamily: dimethylallyltranstransferase

Query Match 46.1%; Score 695.5; DB 2; Length 299;
Best Local Similarity 52.4%; Pred. No. 1.4e-47;
Matches 154; Conservative 35; Mismatches 96; Indels 9; Gaps 3;

y 8 LTVQCRVERALD---ARLPAENILPQTLHQAMRYSVLNGGKTRPLLTATGQALGLPE 64
b 7 LEACVQANQALSRRFIAPLPQN---TPVETWYQVALLGKRLPFLVYATGEMFGVST 63
y 65 NVLDAPACAVEFHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 124
b 64 NTLADAPAAVECHAYSLIHDDLPAMDNDLRRGKPTCHKVGEANAILAGDALQALAF 123
y 125 VLANDPGITVDAPARLKMITALTRASSGQMGVGGQIDLGSGVRKLTLPLEENMH 184
b 124 IESDADMPVSDRDRISMISLASGIAGNCGGQALDLDAGKGVPLDALARIRHRTG 183
y 185 ALIRASVNLAAALSKPDLDTCAKLDHYAKCIGLSFQVNDLIDLEADTATLGTQKDI 244
b 184 ALIRAAVRLGALGAGDKGRRALPVLDKYAESIGLAFQVQDDILDVGTATLGTQKQADQ 243
y 245 DNDKPTYPALLGWAGAKQAQELHEQAVESLTGFGSEA---DLLRELSLYIERT 295
b 244 QLGKSTYPALLGLEQAKQAQDLIDARQSLQAKAQSDITSALEALADYIIQR 297

RESULT 8
JQ0665
Geranyltranstransferase (EC 2.5.1.10) - Escherichia coli (strain K-12)
;Species: Escherichia coli
;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 01-Mar-2002
;Accession: JQ0665; E64771
;R; Fujisaki, S.; Hara, H.; Nishimura, Y.; Horiuchi, K.; Nishino, T.
J. Biochem. 108, 995-1000, 1990
;Title: Cloning and nucleotide sequence of the ispa gene responsible for farnesyl diphos
;Reference number: JQ0664; MUID:91210228; PMID:2089044
;Accession: JQ0665
;Molecule type: DNA
;Residues: 1-299 <FUJ>
;Cross-references: GB:D00694; NID:9216582; PIDN:BA00599.1; PID:9216584
;Experimental source: strain K-12
;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
```


A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of *Escherichia coli* K-12.
A.Reference number: A64720; MUID:97426617; PMID:9278503
A.Accession: E64771
A>Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-299 <BLAT>
A.Cross-references: GB:AB00148; GB:U00096; NID:g1786614; PIDN:AAC73524.1; PID:g1786623;
A.Experimental source: strain K-12, substrain MG1655
A.Comment: This enzyme catalyzes the condensation of isopentenyl diphosphate with dimethyl
F:84-93/Region: aspartate-rich
F:223-227/Region: DDXD motif
F:241-248/Region: nucleotide-binding motif A (P-loop)

Query Match 45.8%; Score 591.5; DB 2; Length 299;
Best Local Similarity 52.0%; Pred. No. 2.8e-47;
Matches 153; Conservative 36; Mismatches 96; Indels 9; Gaps 3;

QY 8 LTVCOERVERALD---ARLPAENILPOTLHOAMRYSVINGGKTRPLLYATQALGLPE 64
DB 7 LEACVQKQANQSFIAPLFPQN---TPVETWYQALGKRLRPLLYATQALGLPE 63
QY 65 NVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 124
DB 64 NTLDPAAAVECHAYSLIHDDLPAMDNDLRRGLFCHVKGKGNALLAGDALQALAF 123
QY 125 VLNDPGITVDAPARLKMITALTRASSQGMVGQAIIDGSGVRKUTLPLENNHIHTG 184
DB 124 ILSDAMPVSDRDRISMISLASAGTAGCGQALDLAGKHVPLDALERHHRKTG 183
QY 185 ALIRASVNLALSKEPDLTCAKLDHYAKCIGLSFOVKDDILDEADTATLCTGKDI 244
DB 184 ALIRAVRLGALSAGDKRRALPVDKYAESIGLAFVQDDILDVGDATLGRGDAQ 243
QY 245 DNDKPTYPALLGMAGAKQKQAEHQAVESLTGFSRA---DLLRELSYIIE 295
DB 244 QLGRKSTYPALLGKARQKARDLDDARQSLKOLAEQSLDTSALEADYIIOR 297

RESULT 9
264123
geranyltransferase (EC 2.5.1.10) - Haemophilus influenzae (strain Rd K920)
A.Species: Haemophilus influenzae
A.Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
A.Accession: C64123
A.Authors: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirsnes, E.P.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Lieb, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A.Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A.Reference number: A64000; MUID:95350630; PMID:7542800
A.Accession: C64123
A>Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-295 <TIGR>
A.Cross-references: GB:U32822; GB:U32823; NID:g1574265; PIDN:AAC23087.1; PID:g1574277; F:241-248/Region: nucleotide-binding motif A (P-loop)
A.Genetic: ispa
A.Function: catalyzes condensation of isopentenyl diphosphate with dimethylallyl diphosphate
A.Pathway: isoprenoid biosynthesis
A.Superfamily: dimethylallyltransferase
A.Keywords: isoprenoid biosynthesis; transferase
A:85-98/Region: aspartate-rich

F:223-227/Region: DDXD motif

Query Match 44.8%; Score 677; DB 2; Length 295;
Best Local Similarity 49.0%; Pred. No. 3.9e-46;
Matches 145; Conservative 50; Mismatches 97; Indels 4; Gaps 3;

QY 1 MSKLKAVITVCOERVERALDARLPAENILPOTLHOAMRYSVINGGKTRPLLYATQAL 60
DB 1 MGHFSEELQVQVTEINFLRQAQFEGIESHNAFLKAMKYALLGKGRVPLVYATQOML 60
QY 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQA 120
DB 61 GAERQTLDYARAAATEAHAYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQS 120
QY 121 LAPEVLNDPGITVDAPARLKMITALTRASSQGMVGQAIIDGSGVRKUTLPLENNHI 180
DB 121 FAFELITKTPNISTE--QKALIQILAQAGAGVQCMCLGSLDLTISEHKQISLSLEIHR 178
QY 181 HKTGALIRASVNLALSKEPDL-OTCAKLDHYAKCIGLSFOVKDDILDEADTATLGT 239
DB 179 NKTGALLIALKLGKFCSPHPTDKRLQESITQYAAIGLAFVQDDILDEGDSASIGKQ 238
QY 240 QGKXIDNDKPTYPALLGMAGAKQKQAEHQAVESLTGFSRA---DLLRELSYIIE 295
DB 239 VGADLDLKDSTYPLGLSGAKQKQADLYQSALSSELEKIPEDT-TVRALAEFIITR 293

RESULT 10
D82778
geranyltransferase (farnesyl-diphosphate synthase) XF0661 [imported] - Xylella fastidiosa
A.Species: Xylella fastidiosa
A.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
A.Accession: D82778
A.Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
A.Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A.Reference number: A82515; MUID:20365717; PMID:10910347
A.Note: for a complete list of authors see reference number A59328 below
A.Accession: D82778
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-291 <STM>
A.Cross-references: GB:AB003910; GB:AB003849; NID:g9105532; PIDN:AAF83471.1; GSPDB:GN00
A.Experimental source: strain 9a5c
A.Authors: Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, I. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
A.Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.B.; Kuramae, E.E.; Laig chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, I. A.; Authors: Martins, E.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y. ; P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Savasal A.; Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; A.Reference number: A59328
A.Contents: annotation
A.Genetics:
A.Gene: XF0661
A.Superfamily: dimethylallyltransferase

Query Match 44.7%; Score 675; DB 2; Length 291;
Best Local Similarity 50.2%; Pred. No. 5.6e-46;
Matches 149; Conservative 37; Mismatches 99; Indels 12; Gaps 3;

QY 5 KAVITVCOERVERALDARLPAENILPOTLHOAMRYSVINGGKTRPLLYATQALGLPE 64
DB 3 BALTRWRQRTDSVLETLTISPTCAQPLRHAAMRYTVLSGKIRPLLYVYAGHLFVDS 62
QY 65 NVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 124
DB 63 PLLDVPAAAVELIHAYSLVHDDLPAMDNDLRRGRPTVTHAFDEATAILGDTALQALAF 122

235 GXDAIQQKSTFPALLSVEGAKCYLQELAERLYTQLHPYGERAAPLTALARLAVERAH 291

Accession 11
781217
peranlystransferase NMB0261 [imported] - *Neisseria meningitidis* (strain MC58 serogroup 14)
Species: *Neisseria meningitidis*
Date: 31-Mar-2000 #sequence revision 31-Mar-2000 #text change 19-Jan-2001

Accession: F81217
 Tettstein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.

Cite: Science 287, 1809-1815, 2000
Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappapoli, R.; Vignelli, A.; Masignani, V.; Masignani, V.; Scariato, V.; Gill, J.; Wamathavan, J.; Gill, J.
Title: Complete genome sequence of Newcastle disease virus

;Accession: F81217
 ;Reference number: A81090; MUID:2015755; PMID:10710307
 ;Complete genome sequence of Neisseria meningitidis serogroup B strain MC36.
 ;Status: preliminary

```

: Molecule type: DNA
: Residues: 1-259 <TET>
: Cross-references: GB:AE002383; GB:AE002098; NID:G7225484; PIDW:AAF40715.1; PID:G7225484

```

Experimental source: serogroup B, strain MC58
Genetics:
Gene: NME0261

```

;superfamily: dimethylylalanine transferase
Query Match 44.2%; Score 668; DB 2; Length 259;
Best Local Similarity 52.5%; Score 762; DB 2; Length 259;
NCBI 259;

```

37 MRYSLVNGGKGRTPLLTYATGQALGLPENVLDPACAVEFIHVYSLIHDDLPAMNDOLR 96
Matches 137; Conservative 39; Mismatches 83; Indels 2; Gaps 1;

1 MRYAALDGGKRLRPMVLVLAASELGEAVHEAVEQAAAEIMHYISLVHDDMPANDDSLR 60

97 RGKPTCHKAYDEATAI LAGDQALQALAFVLPNDPGITVDAPARLXMITALTTRAGSQGMV 156
61 RGKPTCHIKYGEATALLTGDAQTQAFDVLSPR--TELPAAQLAMLSVLKAGGSRGWA 118

157 GGQAIDLGSVGRKLTLPLELNMIHKTCALIRASVNLAAISKPDLDTCVAKKLHLYAKCI 216

119 GGQAIDLANGKQMCADLEOMHSIKTCALIRAAVILGATACPDIDSAKISVLDAVAAKI 178

217 GLSFQKDDILDIEDNTATLTKGTQGGKOIDNDKPTYPALLGNAGAKQKQBELHEQAVESLT 276

179 GLAQVIDDVLDCRBDATFLGKTAGCADNDKPTTVVKLMGLEAARSYAHKLVAEAVALLE 238

277 GFGSEADLLISELSLYIERETH 297

b ||| : ||| : ||| : ||| :
239 PFGDKALRLQLAEFAVARKY 259

ESULT 12
eranylgeranyl pyrophosphate synthase - *Synechocystis* sp. (strain PCC 6803)
74538

;Alternate names: hypothetical protein sir0739
;Species: *Synechocystis* sp.
;Variety: PCC 6803

S.; Kotani, M.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Chikama, H.; Koushi, Y.; Kato, T.; Sato, T.; Ohtsuka, S.; Okumura, T.; Kaneko, T.; Sato, T.; Date, 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

1. K.; Oshimura, S.; Shimo, S.; Takeuchi, C.; Wata, I.; Watanabe, A.; Yamada, M.; Issuda, NA Res. 3, 109-136, 1996

Db 196 LASATSSDVSPVRVAVRAGLAKAIGTEGVAGQVVDISEGDLNDVGLHLEFHLH 255
QY 182 KTGALIRASVNLAAALSKPDLDTVCVAKLDHYAKICIGLSFQVKDDILOEADTATLGTQ 241
Db 256 KTAALLAESAVLGAIVGGSDDEI--ERLRKFCARCIGLLFQVVDILDTVTKSSKELGTAG 314
QY 242 KDIDNDKPTYPALLGMAGAKAKOAEHQAVESLTGFGSE--ADLLRELSLYIIER 295
Db 315 KDLIADKLTYPKIMGLEKSRFAEKLNRDQLLGFDSKVAPLL-ALANYIAYR 369

RESULT 14
S53722
farnesyltransferase (EC 2.5.1.29) precursor - pepper
N:Alternate names: geranylgeranyl diphosphate synthase; geranylgeranyl pyrophosphate syn
C:Species: capsicum annuum (pepper)
C:Date: 15-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
C:Accession: S53722
R:Badillo, A.; Steppuhn, J.; Deruere, J.; Canara, B.; Kuntz, M.
Plant Mol. Biol. 27, 425-428, 1995
A:Title: Structure of a functional geranylgeranyl pyrophosphate synthase gene from Capsi
A:Reference number: S53722; MUID:95195169; PMID:7888631
A:Accession: S53722
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <BAD>
A:Cross-references: EMBL:X80267; NID:G643093; PIDN:CAAS6554.1; PID:G643094
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Genetics:
A:Gene: GGPPS
A:Genome: nuclear
C:Superfamily: geranyltransferase
C:Keywords: carotenoid biosynthesis; chloroplast; chromoplast; transferase

Query Match 39.5%; Score 596; DB 2; Length 369;
Best Local Similarity 47.1%; Pred. No. 1.3e-39;
Matches 139; Conservative 45; Mismatches 101; Indels 10; Gaps 7;

QY 5 KAYLTVCQERVERALDARLPAENILPQTLHOAMRYSVLNGGKRTPLLTATGQALGLPE 64
Db 79 KIYVTEKAIKSNKALDEALIVKE--PHVTHEAMRYSLLAGGKVRPMLCLAACELVGENQ 136
QY 65 NVLDAPACAVEFIHYVSLIHDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQALAFE 124
Db 137 ENAMAACAACAVEMIHMTSLIHDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQALAFE 196
QY 125 VLAND-PGIVTDAPARL-KMITALTFRASGQGMVGQAIDLGVSGR-KLTLPLENNMHIH 181
Db 197 HIVNSTAGVT---PSRIVGAVAEALAKS:GTEGLVAGQVADIKCTGNASVSLETLEFIHVH 253
QY 182 KTGALIRASVNLAAALSKPDLDTVCVAKLDHYAKICIGLSFQVKDDILOEADTATLGTQ 241
Db 254 KTAALLLESSVLGAILGGTNEV-EKLRKFCARCIGLLFQVVDILDTVTKSSSEELGTAG 312
QY 242 KDIDNDKPTYPALLGMAGAKAKOAEHQAVESLTGFGS--RADLLRELSLYIIER 295
Db 313 KOLVNDKTTYPKIMGLEKSRFAEKLNRDQLLGFDSKVAPLL-ALANYIAYR 367

RESULT 15
T10452
farnesyltransferase (EC 2.5.1.29) precursor, chloroplast - white mustard
N:Alternate names: geranylgeranyl-diphosphate synthase; geranylgeranyl-pyrophosphate syn
C:Species: Sinapis alba (white mustard)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10452
R:Bonk, M.; Hoffmann, B.; von Lintig, J.; Schledz, M.; Al-Babili, S.; Hobeika, E.; Klein
Eur. J. Biochem. 247, 942-950, 1997
A:Title: Chloroplast import of four carotenoid biosynthetic enzymes in vitro reveals dif
A:Reference number: Z17023; MUID:97433278; PMID:9286918
A:Accession: T10452
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-366 <BON>
A:Cross-references: EMBL:X90795; NID:G1419757; PIDN:CA67330.1; PID:G1419758
C:Genetics:
A:Gene: GGPS
A:Genome: nuclear
C:Superfamily: geranyltransferase
C:Keywords: carotenoid biosynthesis, chloroplast, transferase

Query Match 39.1%; Score 590.5; DB 2; Length 366;
Best Local Similarity 45.6%; Pred. No. 3.6e-39;
Matches 135; Conservative 52; Mismatches 98; Indels 11; Gaps 7;

QY 6 AYLTVCQERVERALDARLPAENILPQTLHOAMRYSVLNGGKRTPLLTATGQALGLPEN 65
Db 74 SYIIRKADSVNKALDSAVPLRE--PLKIHEAMRYSLLAGGKVRPVLICIAACELVGGES 131
QY 66 VLDAPACAVEFIHYVSLIHDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQALAFEV 125
Db 132 LAMPARCAVEMIHMTSLIHDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQALAFEV 191
QY 126 LANDPGITVDAPAR-LKMITALTFRASGQGMVGQAIDLGVSGRKLT---LPELENMHIH 181
Db 192 LASATSSSEV-SPARVVRVAVRAGLAKAIGTEGLVAGQVVDISEGDLNDVGLHLEFHLH 250
QY 182 KTGALIRASVNLAAALSKPDLDTVCVAKLDHYAKICIGLSFQVKDDILOEADTATLGTQ 241
Db 251 KTAALLAESAVLGAIVGGSDDEI--ERLRKFCARCIGLLFQVVDILDTVTKSSBLGKTAG 309
QY 242 KDIDNDKPTYPALLGMAGAKAKOAEHQAVESLTGFGSE--ADLLRELSLYIIER 295
Db 310 KDLIADKLTYPKIMGLEKSRFAEKLNRDQLLGFDSKVAPLL-ALANYIAYR 364

Search completed: February 29, 2004, 14:52:36
Job time : 11.2669 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:27:18 ; Search time 5.00562 Seconds
(without alignments)
3089.496 Million cell updates/sec

title: US-09-941-947A-20

effect score: 1510

sequence: 1 MSKIKAYLVCQVERALD.....FGSEADLLRELSLYIIERTH 297

coring table: BLOSUM62

Gapop 10.0 , Capext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	691.5	45.8	299	1 ISPA_ECOLI	P22939 escherichia
2	677	44.8	295	1 ISPA_HAEN	P45204 haemophilus
3	604	40.0	371	1 GGPP_ARATH	P34802 arabidopsis
4	596	39.5	369	1 GGPP_CAPAN	P80042 capsicum an
5	590.5	39.1	366	1 GGPP_SINAL	Q43133 sinapis alb
6	583	38.6	332	1 ISPA_RHISN	P55339 rhizobium s
7	578	38.3	332	1 ISPA_BRAJA	Q45220 bradyrhizob
8	577	38.2	297	1 ISPA_BACST	Q08291 bacillus st
9	575	38.1	357	1 GGPP_CATRO	Q42698 catharanthu
10	566	37.5	291	1 ISPA_MICLU	O66126 micrococcus
11	531	35.2	272	1 ISPA_BACSU	P54383 bacillus su
12	522.5	34.6	300	1 CRTE_CVAPA	P48368 cyanophora
13	501.5	33.2	294	1 ISPA_BUCAP	Q8K9A0 buchnera ap
14	501	33.2	282	1 ISPA_BUCAL	P57537 buchnera ap
15	399	26.4	289	1 CRTE_RHOCA	P17060 rhodobacter
16	369	24.4	262	1 ISPA_AQUAE	O66952 aquifex aeo
17	367	24.3	288	1 CRTE_RHOSH	P54976 rhodobacter
18	367	24.3	325	1 IDSA_METJA	O26156 methanobact
19	353.5	23.4	327	1 IDSA_METJA	Q58270 methanococc
20	353	23.4	324	1 IDSA_METTM	O53479 methanobact
21	336	22.3	302	1 CRTE_PANAN	P21684 partoea ana
22	317.5	21.0	307	1 CRTE_ERWHE	P22873 erwinia her
23	290.5	19.2	323	1 ISPB_ECOLI	P19641 escherichia
24	290.5	19.2	332	1 GGPP_SULAC	P95999 sulfolobus
25	282	18.7	330	1 GGPP_SULAC	P39464 sulfolobus
26	281	18.6	329	1 ISPB_HAEN	P44916 haemophilus
27	277.5	18.4	323	1 PREA_SVNY3	P72580 synechocyst
28	270	17.9	359	1 GGPP_XYCTU	Q50727 m probable
29	265	17.5	323	1 PREA_PORPU	P51268 porphyra pu
30	259	17.2	323	1 PREA_CTAPA	P31171 cyanophora
31	244	16.2	320	1 HEP2_BACST	P55785 bacillus st
32	226	15.0	348	1 HEP2_BACSU	P31114 bacillus su
33	215.5	14.3	378	1 DPS_SCHPO	Q43091 schizosacch

ALIGNMENTS

RESULT 1

ID	ISPA_ECOLI	STANDARD;	PRT;	299 AA.
AC	P22939;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Geranyltransferase [EC 2.5.1.10] (Farnesyl-diphosphate synthase)			
DE	(ppp synthase)			
GN	ISPA OR B0421.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KL2;			
RC	MEDLINE=S1210228; PubMed=2089044;			
RA	Fujisaki S., Hara H., Nishimura Y., Horiuchi K., Nishino T.;			
RT	"Cloning and nucleotide sequence of the ispa gene responsible for			
RT	farnesyl diphosphate synthase activity in Escherichia coli.";			
RL	J. Biochem. 108:995-1000(1990)			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KL2;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shaq Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RL	Science 277:1233-1238(1997)			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,			
RA	Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,			
RA	Lew H., Lin D., Nacath A., Oefner P., Schramm S., Davis R.W.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
CC	- - CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate			
CC	= diphosphate + trans,trans-farnesyl diphosphate.			
CC	- - SUBCELLULAR LOCATION: Cytoplasmic.			
CC	- - SIMILARITY: Belongs to the FPP/GGPP synthetase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D00694; BAA00559.1; -			
DR	EMBL; AS000146; AAC73524.1; -			
DR	EMBL; U82664; AAB40177.1; -			
DR	PIR; JQ0665; JQ0665.			
DR	SWISS-2DPAGE; P22939; COLI.			

Q9t1s1 cyanidium c
P18900 saccharomyc
Q9wn0 m geranylge
P56966 b geranylge
Q95749 h geranylge
Q92236 g geranylge
P24322 n geranylge
Q43315 arabidopsis
O24242 parthenium
Q09152 arabidopsis
O14230 schizosacch
P49353 zea mays (m

Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P., Langham S.-A., McCullagh B., Bihlham R., Robben J., Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E., Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W., Moolijm F., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Berneser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H., De Keyser A., Buyshaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R., Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S., Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H., Dose S., de Haan M., Maare A., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fartmann B., Granderath K., Daurer D., Herzl A., Neumann S., Fugirion A., Vitale D., Liguori R., Piravandi E., Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R., Schnabl S., Hiller R., Schmidt W., Lecharry A., Aubourg S., Chafdor F., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T., Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C., Frishman D., Haase D., Lencke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., Senkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Krawiec J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W., Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C., Antoniou B., Zidanic M., Strong C., Sun H., Lanar B., Yordan C., Ma P.-K., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shan R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S., Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A., Chen E., Marra M.A., Martienssen R., McCombie W.R.; "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana"; Nature 402:769-777(1999).

-!- FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.

-!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.

-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans-trans-farnesyl diphosphate.

-!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.

-!- PATHWAY: First committed step in carotenoid biosynthesis. Key enzyme in plant terpenoid biosynthesis.

-!- SUBUNIT: Monomer.

-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L25813; AAA32797.1; -
EMBL; Z95708; CAB16803.1; -
EMBL; AL161590; CAB80347.1; -
PIR; F85434; F85434.
InterPro: IPR000092; Polyprenyl synt.
InterPro: IPR008949; Terpenoid synth.
Pfam: PF00348; polyprenyl synt; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
Tranferase; Isoprene biosynthesis; Carotenoid biosynthesis.

KW	Chloroplast; Transit peptide.	1	?	CHLOROPLAST.
FT	CHAIN	?	371	GERANYLGERANYL PYROPHOSPHATE SYNTHETASE.
FT	CONFLICT	108	108	R -> S (IN REF. 1).
FT	CONFLICT	141	141	A -> R (IN REF. 1).
FT	CONFLICT	192	192	A -> S (IN REF. 1).
SEQ	SEQUENCE	371 AA;	40174 MW;	EFA8088A75B6A005 CRC64;

Query Match	40.0%;	Score 604;	DB 1;	Length 371;
Best Local Similarity	45.9%;	Pred. No. 7.9e-40;		
Matches 136;	Conservative 51;	Mismatches 99;	Indels 10;	Gaps 6

QY	6	AYLTVCQERVERALDARLPAENILPQTLHOAMEYSVLNGKTRPLITVATGQALGLPEN	65
DB	78	SYLTITKASLVNKALDSVPLR--PLKHHEAMYSLLAGGRVPLCTARCELVGGBES	135
QY	66	VLDAPACAVEFIHYVSLIHDDLPAONDLLRRGKPTCHKAYDEATAILAGDALQALAPEV	125
DB	136	TAMPAACAVEMHTMSLIHDDLPCMONDDLRRGKPTNHKVFGEVAVLAGDALISPAFEH	195
QY	126	LANDPGIIVDAPAR-LKKVITALTASSQMGWGQALIDGSGVRKLT---LPELNWHIH	181
DB	196	LASATSPDWSVPRVWRVAVGELAKAIGTEGLVAGQVVDISSEGLDLNDVGLHLEHFIHL	255
QY	182	KTGALIRASVNLAAISKPDLDVFCVAKKLDHYAKICIGLSFOVKDDILDIEADTATLGKQG	241
DB	256	KTAALLEASVLAGVAGGSDDEI-ELRKFARFCIGLLFQVVDILDVTKESKELKTAG	314
QY	242	KDINDKPTYPALLGMAGAKOKAQELHQBVESLTGSGE--ADLLRELSLYIIR	295
DB	315	KDLTADKLITYPKIMGLBKSREFAKLNREARDQLLGFDSDKVPALL-ALANVIAYR	369

RESULT 4	GGPP-CAPAN	STANDARD;	PRT;	369 AA.
ID	GGPP-CAPAN			
AC	P80042;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP synthetase) [GGPS] [includes: Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)]			
GN	GGPS1.			
OS	Capsicum annuum (Bell pepper).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;			
OC	Lamiales; Solanales; Solanaceae; Capsicum.			
OX	NCBI_TaxID=4072;			
GN	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=cv. Lamuyo; TIGSUE=fruit;			
RC	MEDLINE=93272043; PubMed=1303794;			
RX	Kuntz M., Roemer S., Suire C., Huguency P., Weil J.H., Schantz R., Camara B.;			
RA	"Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme activity and transcript level during fruit ripening.";			
RT	Plant J. 2:25-34(1992).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=cv. Yolo Wonder;			
RC	MEDLINE=95195169; PubMed=7888631;			
RX	Badrillo A., Steppuhn J., Deruere J., Camara B., Kuntz M.;			
RA	"Structure of a functional geranylgeranyl pyrophosphate synthase gene from Capsicum annuum.";			
RT	Plant Mol. Biol. 27:425-428(1995).			
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				
RC				
RX				
RA				
RT				
RL				
RN				
RP				


```

01-NOV-1997 (Rel. 35, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Probable geranyltransferase [EC 2.5.1.10] {Farnesyl-diphosphate synthase} (PPP synthase).
Y4KU.
Rhizobium sp. (strain NGR234).
Plasmid sym pNGR234a.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
NCBI_TaxID=394;
[1]
SEQUENCE FROM N.A.
MEDLINE=97305956; PubMed=9163424;
Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A., Perret X.;
"Molecular basis of symbiosis between Rhizobium and legumes.";
Nature 387:394-401(1997).
-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diisopentyl + trans-trans-farnesyl diphosphate.
-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential)
-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
EMBL; AE000082; AA931752.1; -.
PIR; T10875; T10875.
InterPro; IPR000092; Polyprenyl synt.
InterPro; IPR008949; Terpenoid_synth.
Pfam; PF00348; polyprenyl synt; 1.
PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
Transferase; Isoprene biosynthesis; Plasmid.
SEQUENCE 332 AA; 34688 MW; EE68C1547449AA87 CRC64;

Query Match      38.68; Score 583; DB 1; Length 332;
Best Local Similarity 45.78; Pred. No. 3e-38;
Matches 133; Conservative 46; Mismatches 98; Indels 14; Gaps 4;

ZY 13 ERVERALDARLPANILPQLTHQAMRYSLVNGKTRPLLTATTCQAAGLGPENVLDPAC 72
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||
Yb 42 KRVEQALARULCABDHCETELMAARYATLHGGKRTRALLCLAAGALADTPAHILDVGA 101
   :|::||::||::||::||::||::||::||::||::||::||::||::||

ZY 73 AVFIHTVSLIHDDLPMNDNDLRGKPTCHKRYDBEATNLAGDLQALAEVLANDPGI 132
   :|::||::||::||::||::||::||::||::||::||::||::||::||
Yb 102 AIENMWHACTLVHDDLPAMDVDVLRGLFTVHKFGBEATAILVGDLAQAHAFULTA ---- 156
   :|::||::||::||::||::||::||::||::||::||::||::||::||

ZY 133 TVDAPA--RLKNMTALTFRASGSQMVGQAIDLGSVGRKLTLPELENHHIKTGALIRAS 190
   :|::||::||::||::||::||::||::||::||::||::||::||::||
Yb 157 SLDPAGDNRRALTVELLAQAVSAGSAGGQAMDLSLVGEHVLDRIVANHRVKCALVRAS 216

ZY 191 VN---LAALSXPOLDTCVAKKLDRHYAKCIGLSFQVKDDIIDLEADTATLGKTQGDIND 247
   :|::||::||::||::||::||::||::||::||::||::||::||::||
Yb 217 VRMGALCAIAEDAADATLYLCALDFYSACFGLAQVVDDILDATATLGKTQGDAAAQ 276
   :|::||::||::||::||::||::||::||::||::||::||::||::||

ZY 248 KPTYPALGLHWAGAKOKAQELHEQAVESLTGFSEAD----LLRELSTVIIE 294
   :|::||::||::||::||::||::||::||::||::||::||::||::||
Yb 277 KPTCASIMGLOANQFALDLLCSAGEATAPLGPRAEIRLAQMQLQRASYIFK 327
   :|::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 7
ID ISPA BRAJA
IC Q45220; STANDARD; PRT; 332 AA.
YI 01-NOV-1997 (Rel. 35, Created)
YI 01-NOV-1997 (Rel. 35, Last sequence update)
YI 10-OCT-2003 (Rel. 42, Last annotation update)
Probable geranyltransferase [EC 2.5.1.10] {Farnesyl-diphosphate synthase} (PPP synthase).

```

synthase) (FPP synthase).
FPS OR BLR2148.
Bradyrhizobium japonicum.
Sbacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
C Bradyrhizobiaceae; Bradyrhizobium.
XNCBL_TaxID=375;
L[1]
P
FSEQUENCE FROM N.A.
ASTRAIN-USDA 110;
Ttully R.E., Keister D.L.;
T"Cloning and mutagenesis of a cytochrome P-450 locus from
TBradyrhizobium japonicum that is expressed anaerobically and
Lsymbolically."; Appl. Environ. Microbiol. 59:4136-4142(1993). [2]

SEQUENCE FROM N.A.
MEDLINE=USDA 110;
MEDLINETitle=98322110; PubMed=9655913;
Tully R.E., van Berkum P., Lovins K.W., Keister D.L.;
Identification and sequencing of a cytochrome P450 gene cluster from
Bradyrhizobium japonicum";
Biochim. Biophys. Acta 1398:243-255(1998). [3]

SEQUENCE FROM N.A.
MEDLINE=USDA 110;
MEDLINE=22484998; PubMed=12597275;
Kaseno T., Nakamura Y., Sato S., Minamisawa K., Uchimi T.,
Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
Kohara M., Matsumoto M., Shimpo S., Tsurukawa H., Wada T., Yamada M.,
Tabata S.;
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
DNA Res. 9:189-197(2002).

-! CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
= diphsophate + trans-trans-farnesyl diphosphate.
-! SUBCELLULAR LOCATION: Cytoplasmic Potential.
-! SIMILARITY: Belongs to the fpp/GGPP synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

EMBL; U12678; AAC28894.1; ..
EMBL; AP2005942; BAC47413.1; ALT_INIT.
PIR; I40213; I40213.
InterPro; IPRO000052; Polyprenyl synt.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synt; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
Transferrase; Isoprene biosynthesis; Complete proteome.
SEQUENCE 332 AA; 3438 MW; 47644FAF0220CEAC CRC64;

Query Match 38.3%; Score 578; DB i; Length 332;
Best Local Similarity 47.3%; Pred. No. 7.4e-38;
Matches 133; Conservative 40; Mismatches 98; Indels 10; Gaps 4;

y y 13 EVVEREALDARLPANILPQTLHQAWRYSVLVNGSKRTPEPLTYATGGALGPENVLDPAPC 72
 :|||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
b b 42 KRVEEARLARLICADDEGETELMAAMRWYATVHGGKTRALLCLAGALADTPAHMLDVGA 101
 ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
y y 73 AVFEHYVSLSIHDDLPAWMDNLLRGPKTCCKAYDEATAIAGDALQAALAFEVLA--NDP 130
 ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
b b 102 ALLEMMHACTLVDDDLPAWDDVLRGLPTVVHVKFGEATAILVGDAQAHLFLASLNAP 161
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
y y 131 GITVDAPARKLMITALTFRSGSQGVCGOAILDGSCVKLTLELENMHIHKTGALIRAS 190
 ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
b b 162 G---DSP---IALVELAQAVSAEGAAQGQAIIDLSLVGKHVELDRIVVAHHMKMGALVRAS 216

QY 191 VMLAALSKEDLTCVAK---KLDHYAKICIGISFQVKDDILDEADTATLGTQKQDIND 247
DB 217 VRMGALCAVGNVNAARALYCALDHYSAFGLAQVDDILSVTATLGTQKPKGDAQA 276
QY 248 KTYTPALLGMAKAKQKQAEHQEQAQVESITGFGSEADILREL 288
DB 277 KTCASIMELQARQFALDILLEDAGEAIPGPRALRAQL 317

RESULT 8
ISPA_BACST
ID ISPA_BACST STANDARD; PRT; 297 AA.
AC Q0831; Q53435; Q53436; Q53437; Q53438;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
DE (FPP synthase).
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=ATCC 10149;
RC MEDLINE=93252758; PubMed=8486607;
RA Koyama T., Obata S., Osabe M., Takeshita A., Yokoyama K.,
RA Uchida M., Nishino T., Ogura K.;
RT "Metabolite farnesyl diphosphate synthase of Bacillus
RT stearothermophilus: molecular cloning, sequence determination,
RT overproduction, and purification.";
RL J. Biochem. 113:355-363 (1993).
RN [2]
MUTAGENESIS OF CYSTEINE RESIDUES.
RP MEDLINE=95001990; PubMed=7918490;
RA Koyama T., Obata S., Saito K., Takeshita-Koike A., Ogura K.;
RT "Structural and functional roles of the cysteine residues of Bacillus
RT stearothermophilus farnesyl diphosphate synthase.";
RL Biochemistry 33:12644-12648 (1994).
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
DB EMBL; D13293; BAA02551.1; -
DR EMBL; S72629; AAB32272.1; -
DR EMBL; S72630; AAB32273.2; ALT_SEQ.
DR EMBL; S72633; AAB32274.1; -
DR EMBL; S72635; AAB32275.2; ALT_SEQ.
DR PIR; JX0257; JX0257. Polyrenyl synt.
DR InterPro; IPR000092; Polyrenyl synt.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyrenyl synt; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
KW Transferase; Isoprene biosynthesis.
FT MUTAGEN 73 73 C->F; S: NO LOSS OF ACTIVITY.
FT MUTAGEN 289 289 C->P; S: NO LOSS OF ACTIVITY.
SQ SEQUENCE 297 AA; 32310 MW; 0F921C3F029EEB6 CRC64;
Query Match 38.2%; Score 577; DB 1; Length 297;
Best Local Similarity 44.6%; Pred. No. 7.7e-38;
Matches 133; Conservative 42; Mismatches 114; Indels 8; Gaps 3;
QY 4 LKAYLTCQVERALD---ABLPRNLTQTLHQAEMYSVLNGGKRTRPLLTATGAL 60

DB 6 VEQFLNEQQAQVETALSRYIERLEG----PAKIKCAMAYSLEAGKRIPLLLSTVRAL 61
QY 61 GLPENVLDAQACAVAFINHYSLIHDDLPAMDNDLRRGKPTCHKAYDAFATLADGALQA 120
DB 62 GQDPAVELPVAACIAEMINTYSLIHDDLPAMDNDLRRGKPTNKKVFGEMAILAGDGLUT 121
QY 121 LAFEVLANDPGITVDAPARLKMITALTATRASQSGQGVGGCAIDLAGSVGRKLTLPLENNHI 180
DB 122 YAFQLITEIDDERIPPSVRLRIERLAKAAGPGNVAGQAADMEGKTLTSLSELYIHR 181
QY 181 HMTGALIRASVNLAAKSPDLDTCVAKKLDHYAKICIGISFQVKDDILDEADTATLGTQ 240
DB 182 HKTGMQLQSVHAGALI--GGADARQTRDELDEFAHLGLAFQIRDDILDIEGAEEKIGEPV 240
QY 241 GKDINDKTTYPALGMAKAKQKQAEHQEQAQVESITGFGSEADILRELSELYIERTH 297
DB 241 GSDQGNKATYPALLSLAGAKELAFHIAAQRHLRNADVGGALAYICELVAARDH 297

RESULT 9
GGPP_CATRO
ID_GGPP_CATRO STANDARD; PRT; 357 AA.
AC Q42698;
DT 16-OCT-2001 (Rel. 40, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP
DE synthetase) (GGPS) [Includes: Dimethylallyltransferase (EC 2.5.1.1);
DE Geranyltransferase (EC 2.5.1.10); Farnesyltransferase
DE (EC 2.5.1.29)].
GN GGPS1 OR GGC1.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Sukkoryta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Gentianales; Apocynaceae; Rauvolfioideae; Vinceae;
OC Catharanthus.
OX NCBI_TaxID=4058;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. G. Don c20;
RA Bantignies B., Liboz T., Ambid C.;
RT "Nucleotide sequence of a Catharanthus roseus geranylgeranyl
RT pyrophosphate synthase gene.";
RL (in) Plant Gene Register PGR95-119.
CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of
CC IPP onto DMAPP to form geranylgeranyl pyrophosphate.
CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key
CC enzyme in plant terpenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
DB EMBL; X32893; CAA63486.1; -
DR PIR; T09966; T09966.
DR InterPro; IPR000092; Polyrenyl synt.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyrenyl synt; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
KW Transferase; Isoprene biosynthesis; Carotenoid biosynthesis;

Chloroplast; Transit peptide. CHLOROPLAST (POTENTIAL).
TRANSIT 1 40
CHAIN 41 357 GERANYLGERANYL PYROPHOSPHATE SYNTHETASE.
SEQUENCE 357 AA; 38786 MW; 81C52FDDAL806FAB CRC64;
Query Match 38.1%; Score 575; DB 1; Length 357;
Best Local Similarity 47.1%; Pred. No. 1.4e-37;
Matches 139; Conservative 43; Mismatches 103; Indels 10; Gaps 7;
Y 5 KAVITVQCEVERAL-DARLPAENILPQTHQAMRYSLVNGKRTPLITYATGOALGUP 63
b 67 KAYIGKANSNNKALDEAVLVRE---PLKHESMYSLLAGGRVPMUCIAACELFGGT 123
Y 64 ENVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATALLAGDALQALAF 123
b 124 ESVAMPACAVEMHITWSLMHDDLPAMDNDLRRGKPTCHKAYDEATALLAGDALQALAF 183
Y 124 EVLAN-DPGITVDAPARLKMITALTRASSQGVGGQADLGSVG-RKLTPELENMHIH 181
b 184 EHTATATKGVSSERIVR--VVGELAKICIGSEGVAGQVVDVCSBGEIADVGLHLEFIH 241
Y 182 KTGALTRASVNLAAKSPDLDTCAVKLDHYAKICIGLSFQVKKDILDIADTATLQKTOG 241
b 242 KTAALLEGSVLGAIVGANDDEIS-KLRFKARCIGLLFQVDDILDVTKSSQELGKTAG 300
Y 242 KQDNDKPTYPALLGMAGAKQAKQELHQAESLTGFGSE-ADLRLRELSLYIER 295
b 301 KDLVADKVTYKLLGIDKREFAKLNRAEQLEAFDPEKAAPLIALANYIAYR 355
RESULT 10
SPA_MICLUJ STANDARD; PRT; 291 AA.
C 066126;
T 15-JUL-1999 (Rel. 38, Created)
T 15-JUL-1999 (Rel. 38, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Geranyltransterase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
F (FPP synthase).
N FPS.
S Micrococcus luteus (Micrococcus lysodeikticus).
C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
C Micrococineae; Micrococcaceae; Micrococcus.
X NCBI_TaxID=1270;
N [1]
P SEQUENCE FROM N.A.
X STRAIN=B-P 26;
X MEDLINE=98175686; PubMed=9515931;
A Shimizu N., Koyama T., Ogura K.;
T "Molecular cloning, expression, and characterization of the genes
encoding the two essential protein components of Micrococcus luteus
B-P 26 hexaprenyl diphosphate synthase.";
T J. Bacteriol. 180:1578-1581 (1998).
L C.-I. CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
= diphosphate + trans,trans-farnesyl diphosphate.
C -I- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
C -I- SIMILARITY: Belongs to the FPP/G3PP synthetase family.
C This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
C EMBL; AB003187; BAA25265.1;
R InterPro; IPR000092; Polyprenyl synth.
R InterPro; IPR008949; Terpenoid synth.
R Pfam; PF00348; polyprenyl synth.
R PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
R PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
N Transferase; Isoprene biosynthesis.

SQ SEQUENCE 291 AA; 32362 MW; 797125AB71B5674A CRC64;
Query Match 37.5%; Score 566; DB 1; Length 291;
Best Local Similarity 45.5%; Pred. No. 5.4e-37;
Matches 130; Conservative 44; Mismatches 92; Indels 20; Gaps 5;
QY 16 ERAIDARLPAENILPQTHQAMRYSLVNGKRTPLITYATGOAL-----GLPENVLDA 69
b 18 ESLNKVHPAQs---RLHEAINVSLSAGGKRPLJLVLTLSLGSNAHDGLPFGI--- 70
QY 70 PACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATALLAGDALQALAFVLAN 129
b 71 ---ALEMIHTYSLIHDDLPAMDNDLRRGKPTCHKAYDEATALLAGDALQALAFVLAN 126
QY 130 PGITVDAPARLKMITALTRASSQGVGGQADLGSVG-RKLTPELENMHIHKTGELIRA 189
b 127 --TQLNNEIKSLINLLSTAGSGVGGVGMQMDQGEHKTLINELERHIHKTGELIRA 184
QY 190 SVNLAALSKPDLPTCAVKLDHYAKICIGLSFQVKKDILDIADTATLQKTOGKIDNDKP 249
b 185 AIVSAGIIM-NFNDAQTEQLNIIGKNVGLMFQIKDDILDVEGSEFENIGKTVGSDLNDRS 243
QY 250 TYPALLGMAGAKQAKQELHQAESLTGFGSEADLRLRELSLYIER 295
b 244 TYVELLGLEASKOLLNDKLTETTYDAUKTLQPINDKLTITTYIVER 289
RESULT 11
ISPA_BACSU STANDARD; PRT; 272 AA.
AC P54383;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranyltransterase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
DS (FPP synthase).
GN YQID OR BSU24289.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8959508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
the Bacillus subtilis genome containing the skin element and many
sporulation genes.";
RL Microbiology 142:3103-3111 (1996).
RP [2]
P SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignall S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.F.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,


```

RESULT 13
ID ISPA_BUCAP STANDARD; PRT; 294 AA.
AC Q8K9A0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
DE (FPP synthase).
EN ISPA OR BUSG449.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbeek B., Naeslund A.K., Eriksson A.-S.,
RT Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.R.;
RL "50 million years of genomic stasis in endosymbiotic bacteria.";
Science 236:2376-2379(2002).
CC -!- CATABOLIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the fpp/gpp synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB014121; AAMG7992.1; -.
CC InterPro; IPR000092; Polyprenyl synth.
CC InterPro; IPR008949; Terpenoid synth.
CC Pfam; PF00348; Polyprenyl synth; 1.
CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
CC Transferrase; Isoprene biosynthesis; Complete proteome.
CW SEQUENCE 294 AA; 33880 MW; 7D7D901C7213ED0A CRC64;
SQ
Query Match 33.2%; Score 501.5; DB 1; Length 294;
Best Local Similarity 42.2%; Pred. No. 5.9e-32;
Matches 117; Conservative 49; Mismatches 98; Indels 13; Gaps 5;
2Y 23 LPAENILPQTLHQMRYSVLNGKRTPLITYATGQALGPLPENVLDPACAVEFIHVSYL 82
DB 25 LPFOD---SVLFRAMKYSTLSGGKRLRACLIYATGETFQVNI AALDVI SAAVELHVSYL 81
2Y 83 IHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN--PGITVDAPARL 140
DB 82 IHDLPICDNDYFRGKISKICHKGFENFALLAGDALQGLAFNLSNMPG--VHDSIRL 139
2Y 141 KMTALTTRASSQGWGQALDLSGVGRKLTPELENMHIHKTGALIRASVNLAALESKD 200
DB 140 KMIAEFSNAIGSCMGICQMLDLKERRKINI SELEKINLYKTAFIRCSIRLAYFASN 199
2Y 201 LDTCAVKKLHAKYKIGLSGVQKDLIDLEADTATLTKTGQKIDNDKPTYPALLQWAGA 260
DB 200 FSKVELFILDKFSVIGLAFQIQDILDLKNDIKKLESKKNK----TKNTYPLLIGLKKS 255
2Y 261 KKAQELHEQAVESLTGF--GSEADLLRELSYIIER 295
DB 256 KIKIKELYKEAFFTEILKKNFNVLKLTQFIMKR 292
RESULT 14
ID ISPA_BUCAP STANDARD; PRT; 282 AA.
AC Q8K9A0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
DE (FPP synthase).
EN ISPA OR BUSG449.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbeek B., Naeslund A.K., Eriksson A.-S.,
RT Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.R.;
RL "50 million years of genomic stasis in endosymbiotic bacteria.";
Science 236:2376-2379(2002).
CC -!- CATABOLIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the fpp/gpp synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB014121; AAMG7992.1; -.
CC InterPro; IPR000092; Polyprenyl synth.
CC InterPro; IPR008949; Terpenoid synth.
CC Pfam; PF00348; Polyprenyl synth; 1.
CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
CC Transferrase; Isoprene biosynthesis; Complete proteome.
CW SEQUENCE 282 AA; 32331 MW; 228B31A6BDCD6445 CRC64;
SQ
Query Match 33.2%; Score 501; DB 1; Length 282;
Best Local Similarity 41.9%; Pred. No. 6e-32;
Matches 112; Conservative 56; Mismatches 91; Indels 8; Gaps 4;
QY 13 EREVERALDARL-PAENTILP---QTLHQMRYSVLNGKRTPLITYATGQALGPLPENVL 68
DB 8 ERYKHRIHQKLFYTLNQLPFQKSLKAMKYSVFSGSKRLRSLSTYSTGDVFKVNIITLD 67
QY 69 APACAVEFIHVSYLHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN 128
DB 68 VISTAIFHSYLSIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN 127
QY 129 PGITVDAPARLKMITALTRASSQGWGQALDLSGVGRKLTPELENMHIHKTGALIR 188
DB 128 SFMENVSNLKRKIMSELSYISGSGCMQGNLDLEAKKDVNLSELEIINLYKTSFLMR 187
QY 189 ASVNLAALESKDPLDTCVAKKLDHAKYKIGLSGVQKDLIDLEADTATLTKTGQKIDNDK 248
DB 188 SAVELVVFSSNNFSSKSLTSLDLPFSISIGLAFQIQDILDLKNDIKKLESKKNK---K 243
QY 249 PTYPALLQWAGAKAKOQELHEQAVESL 275
DB 244 HTYPLIIGLDESRRKKIKQLHKHKSFLAL 270
RESULT 15
ID CRTE_RHOCA STANDARD; PRT; 289 AA.
AC F17060;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: February 29, 2004, 14:51:24 ; Search time 19.7191 Seconds
(without alignments)
3180.293 Million cell updates/sec

Title: US-09-941-947A-20

Perfect score: 1510

Sequence: 1 MSKLYALVQCQVERALD.....FQSEADLLRELSYIERTH 297

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1510	100.0	297	9	US-09-934-903-14
2	1510	100.0	297	9	US-09-934-903-14
3	1510	100.0	297	10	US-09-941-947A-20
4	831	55.0	295	9	US-09-815-242-11971
5	791.5	52.4	291	15	US-10-369-493-8506
6	782	51.8	295	15	US-10-369-493-13999
7	719.5	47.6	294	15	US-10-369-493-4670
8	719.5	47.6	294	15	US-10-369-493-7429
9	715	47.4	291	15	US-10-369-493-15449
10	707	46.8	284	15	US-10-369-493-15817
11	707	46.8	284	15	US-10-369-493-16200
12	705.5	46.7	299	9	US-09-815-242-14084
13	691.5	45.8	299	9	US-09-815-242-10069
14	691.5	45.8	299	15	US-10-369-493-732
15	690	45.7	306	15	US-10-369-493-10187

16 682.5 45.2 296 15 US-10-369-493-480 Sequence 480, App
17 677 44.8 295 9 US-09-815-242-11239 Sequence 11239, A
18 675 44.7 291 15 US-10-369-493-17527 Sequence 17527, A
19 672 44.5 291 15 US-10-369-493-9290 Sequence 9290, Ap
20 671 44.4 256 15 US-10-369-493-9546 Sequence 9546, Ap
21 669.5 44.3 296 15 US-10-369-493-21173 Sequence 21173, A
22 661 43.8 276 15 US-10-369-493-9513 Sequence 9513, Ap
23 617 40.9 292 15 US-10-369-493-10617 Sequence 10617, A
24 610 40.4 369 13 US-10-108-915-18 Sequence 18, Appl
25 608 40.3 290 15 US-10-369-493-20893 Sequence 20893, A
26 606.5 40.2 302 15 US-10-369-493-2608 Sequence 2608, Ap
27 586 38.8 309 15 US-10-369-493-18752 Sequence 18752, A
28 585 38.7 310 15 US-10-369-493-19922 Sequence 19922, A
29 576 38.1 321 15 US-10-369-493-12201 Sequence 12201, A
30 574 38.0 294 15 US-10-369-493-18010 Sequence 18010, A
31 572 37.9 367 13 US-10-108-915-26 Sequence 26, Appl
32 571.5 37.8 287 14 US-10-166-225A-159 Sequence 159, Appl
33 565 37.4 294 15 US-10-369-493-17358 Sequence 17358, A
34 565 37.4 350 13 US-10-108-915-22 Sequence 22, Appl
35 559 37.0 261 15 US-10-369-493-9710 Sequence 9710, Ap
36 552.5 36.6 316 13 US-10-108-915-45 Sequence 45, Appl
37 551 36.5 377 9 US-09-934-778-2 Sequence 2, Appl
38 548.5 36.3 297 15 US-10-369-493-16549 Sequence 16549, A
39 541.5 35.9 294 15 US-10-369-493-16943 Sequence 16943, A
40 538.5 35.7 296 15 US-10-369-493-20569 Sequence 20569, A
41 531 35.2 272 15 US-10-369-493-23201 Sequence 23201, A
42 516.5 34.2 287 9 US-09-925-637-64 Sequence 64, Appl
43 516.5 34.2 287 14 US-10-084-205-64 Sequence 64, Appl
44 512.5 33.9 288 9 US-09-815-242-5239 Sequence 5239, Ap
45 512.5 33.9 293 9 US-09-815-242-12583 Sequence 12583, A

ALIGNMENTS

RESULT 1

US-09-934-903-14
; Sequence 14, Application US/09934903
; Patent No. US20020102690A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, Mattheos
; APPLICANT: Odum, J. Martin
; APPLICANT: Schenzle, Andreas J.
; APPLICANT: No. US20020102690A1ton, Kelley C.
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Rouviere, Pierre
; APPLICANT: Picataggio, Stephen
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
; FILE REFERENCE: CIL646 US NA
; CURRENT APPLICATION NUMBER: US/09/934,903
; PRIOR FILING DATE: 2001-08-22
; PRIOR FILING DATE: 60/229,907
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ORF7
US-09-934-903-14

Query Match 100.0%; Score 1510; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 4.4e-152;
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSKLYALVQCQVERALDARLPAENILPOTLHQAMRYSVLNGKGRTPILLTYATQAL 60
DB 1 MSKLYALVQCQVERALDARLPAENILPOTLHQAMRYSVLNGKGRTPILLTYATQAL 60
QY 61 GLPENLIDAPACAVEFTHWVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQA 120

Db 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATLADGALQA 120
Qy 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGVGGQADLGSVGRKLTLPLENWHI 180
Db 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGVGGQADLGSVGRKLTLPLENWHI 180
Qy 181 HKTGALIRASVNLAAALSKPDLDTVCVAKKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240
Db 181 HKTGALIRASVNLAAALSKPDLDTVCVAKKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240
Qy 241 GKXINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRLSLYIERTH 297
Db 241 GKXINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRLSLYIERTH 297

RESULT 2

US-09-934-868-72
; Sequence 72, Application US/09934868
; Patent No. US20020137190A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, Mattheos
; APPLICANT: Odum, James M
; APPLICANT: Schenzle, Andreas J
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 72
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ORF7 - ISPA
US-09-934-868-72

Query Match 100.0%; Score 1510; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 4.4e-152;
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKLKAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGQAL 60
Db 1 MSKLKAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGQAL 60
Qy 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATLADGALQA 120
Db 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATLADGALQA 120
Qy 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGVGGQADLGSVGRKLTLPLENWHI 180
Db 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGVGGQADLGSVGRKLTLPLENWHI 180
Qy 181 HKTGALIRASVNLAAALSKPDLDTVCVAKKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240
Db 181 HKTGALIRASVNLAAALSKPDLDTVCVAKKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240
Qy 241 GKXINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRLSLYIERTH 297
Db 241 GKXINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRLSLYIERTH 297

RESULT 3

US-09-941-947A-20
; Sequence 20, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.

; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odum, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rovierte, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Methylobionas 16a
US-09-941-947A-20

Query Match 100.0%; Score 1510; DB 10; Length 297;
Best Local Similarity 100.0%; Pred. No. 4.4e-152;
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKLKAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGQAL 60
Db 1 MSKLKAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGQAL 60
Qy 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATLADGALQA 120
Db 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATLADGALQA 120
Qy 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGVGGQADLGSVGRKLTLPLENWHI 180
Db 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGVGGQADLGSVGRKLTLPLENWHI 180
Qy 181 HKTGALIRASVNLAAALSKPDLDTVCVAKKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240
Db 181 HKTGALIRASVNLAAALSKPDLDTVCVAKKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240
Qy 241 GKXINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRLSLYIERTH 297
Db 241 GKXINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRLSLYIERTH 297

RESULT 4

US-09-815-242-11971
; Sequence 11971, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
BEST LOCAL SIMILARITY 60.3%; Pred. No. 8.6e-80;
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 11971
LENGTH: 295

TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
S-09-815-242-11971

Query Match 55.0%; Score 831; DB 9; Length 295;
Best Local Similarity 60.3%; Pred. No. 8.6e-80;
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;

Y 4 LKAYLVCCERVERALDARLPAENILPOTLHOAMRYSVLNGKTRPLITYATGALGLP 63
b 2 IAAVQRCQARVDAALDAUFVARELQLYEAMRYSVWNGKRVRLAYAAACALGGA 61
Y 64 ENVIDAPACAVEPIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATATAGDALQALAF 123
b 62 PQADAAACAVELIHAYSLVHDDLPAMDNDLRRGPTTHRAFDATATAGDALQALAF 121
Y 124 EVLANDPGITVDAPARKMITALTTRASGQGMVGGQAIIDLSVGRKLTLPENLHIIHT 183
b 122 EVLADTRNPQEHAVGLEMLTLRAAGSAGMVGQAIIDLSVGVVALDQAALEVHREHT 181
Y 184 GALIRASVNLAAISKPDLOTCAKLDHYKACIGLSFQVKDDILDIEADTATLGTQCKD 243
b 182 GALIEASVRLGALSGRASPASLAALERTAEALGAFQVQDDILDVESETATLGTQCKD 241
Y 244 INDNKPYPALLGMACAKQKQAEHQAESLTGFGSEADLLRELSLYIER 295
b 242 QAHNKPYTALLGLEAKGVALELRDLAALDGPSPADPLQLARYIVER 293

RESULT 5

S-10-369-493-8506
Sequence 8506, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 8506

LENGTH: 291

TYPE: PRT

ORGANISM: Ralstonia metallidurans

S-10-369-493-8506

Query Match 52.4%; Score 791.5; DB 15; Length 291;
Best Local Similarity 57.8%; Pred. No. 1.3e-75;
Matches 163; Conservative 37; Mismatches 79; Indels 3; Gaps 2;

Y 14 RVREALDARLPAENILPOTLHOAMRYSVLNGKTRPLITYATGALGLPENVLDAPACA 73
b 13 RTEAALAEALPSTDTTPTTLTHAMRYAVLNGKRVRLVHAAGVGVATPEACDAACA 72
Y 74 VEPHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATATAGDALQALAEVFLANDPGIT 133
b 73 VENIHAYSLVHDDMPCKDDDDLRGRPTVHKAYDEATATAGDALQALAEVFLAQTALA 132

QY 134 VDAPARKMITALTTRASGQGMVGGQAIIDLSVGRKLTLPENLHIIHTGALIRASVNL 193
DB 133 --REARKVVAELAVASGSGICMCGQAIIDLVNKKAMTREALEGEMTKTGLIRASVNM 190
QY 194 AALSKEPDLTCVAKLDHYKACIGLSFQVKDDILDIEADTATLGTQCKDINDKPTTFA 253
DB 191 GALTGQALVALDRAAAVGLAFQVQDDILDVTDATLGTAGKDAANDKPTTYS 249
QY 254 LGMAGAKQKQAEHQAESLTGFGSEADLLRELSLYIER 295
DB 250 LGLDAAARELAALQRTDAHEALEGFGTRAGRLAELADLVLR 291

RESULT 6

US-10-369-493-13999
Sequence 13999, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 13999

LENGTH: 295

TYPE: PRT

ORGANISM: Pseudomonas fluorescens

US-10-369-493-13999

Query Match 51.8%; Score 782; DB 15; Length 295;

Best Local Similarity 56.9%; Pred. No. 1.4e-74;
Matches 168; Conservative 36; Mismatches 85; Indels 6; Gaps 4;

QY 4 LKAYLVCCERVERALDARLPAENILPQ--TLHQAMRYSVLNGKTRPLITYATGALG 61
DB 2 IAAVSATSQARVVAALLETFLNAP--LPELARLYENKRYSVWNGKRVRLAYAAACALG 59
QY 62 LPENVLDAPACAVEPIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATATAGDALQAL 121
DB 60 GKAEQANGAACAVELIHAYSLVHDDLPAMDNDLRRGQPTTHKFDCAAILAGDGLQSL 119
QY 122 APEVLANDPGIT--VDAPARKMITALTTRASGQGMVGGQAIIDLSVGRKLTLPENLHII 180
DB 120 AFSALL--DPLSLDSADIRLQVVTALAHAGAGAGMVGQAIIDLSVGLKLDQKALEQMR 178
QY 181 HKTGALIRASVNLAAISKPDLPCTVAKLDHYKACIGLSFQVKDDILDIEADTATLGTQ 240
DB 179 HKTGALIEVSVKLGALASGCAEKDELKSLQTYAALGLAFQVQDDILDVESDTETLGRKQ 238
QY 241 GKDDINDKPTYPALLGMACAKQKQAEHQAESLTGFGSEADLLRELSLYIER 295
DB 239 GADIARDKPTYPALLGLDGDANKAYALELRQALHRLRPFDAAREPLRLARYIVDR 293

RESULT 7

US-10-369-493-4670
Sequence 4670, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 4670
 ; LENGTH: 294
 ; TYPE: PRT
 ; ORGANISM: Burkholderia fungorum
 US-10-369-493-4670

Query Match 47.6%; Score 719.5; DB 15; Length 294;
 Best Local Similarity 53.2%; Pred. No. 6.3e-68;
 Matches 151; Conservative 38; Mismatches 90; Indels 5; Gaps 2;
 QY 13 ERVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGPLPENVLDPAC 72
 DB 15 ERVETALDHYLPGATEPATLHEAMRYAVLGGKRVPLCHAGELTGARAECLDAAA 74
 QY 73 AVEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 132
 DB 75 ALEMIVSYLVHDDPMDDDALRRGKPTVHVKYDEATALLVGDALQSQAFVALTSD--- 131
 QY 133 TVDAPA-RLKMITALTTRASGSGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASV 191
 DB 132 -VLAFAQQALVRELALASGSGVGGQIDLASVGHITLTQLETWHRMKTGALLRAAV 190
 QY 192 NLAALSCKPDLDTCVAKKLDHYAKCIGLSFQVKDDILDIEADTATLTKTQKDIINDKPT 251
 DB 191 RMGALAGETPDAAAMRSIDAYSAAVGLAFQVVDLIDVTTSATLGTAGDKAGDKPT 250
 QY 252 PALLGNAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 295
 DB 251 VSIIGLDASRALAAQLRSDAHAALAPFGARAORLAELADLVNR 294

RESULT 8
 US-10-369-493-7429
 ; Sequence 7429, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianteng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 7429
 ; LENGTH: 294
 ; TYPE: PRT
 ; ORGANISM: Burkholderia cepacia
 US-10-369-493-7429

Query Match 47.6%; Score 719.5; DB 15; Length 294;
 Best Local Similarity 53.2%; Pred. No. 6.3e-68;
 Matches 151; Conservative 38; Mismatches 90; Indels 5; Gaps 2;
 QY 13 ERVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGPLPENVLDPAC 72
 DB 15 ERVETALDHYLPGATEPATLHEAMRYAVLGGKRVPLCHAGELTGARAECLDAAA 74
 QY 73 AVEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 132

Db 75 ALEMIVSYLVHDDPMDDDALRRGKPTVHVKYDEATALLVGDALQSQAFVALTSD--- 131
 QY 133 TVDAPA-RLKMITALTTRASGSGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASV 191
 DB 132 -VLAFAQQALVRELALASGSGVGGQIDLASVGHITLTQLETWHRMKTGALLRAAV 190
 QY 192 NLAALSCKPDLDTCVAKKLDHYAKCIGLSFQVKDDILDIEADTATLTKTQKDIINDKPT 251
 DB 191 RMGALAGETPDAAAMRSIDAYSAAVGLAFQVVDLIDVTTSATLGTAGDKAGDKPT 250
 QY 252 PALLGNAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 295
 DB 251 VSIIGLDASRALAAQLRSDAHAALAPFGARAORLAELADLVNR 294

RESULT 9
 US-10-369-493-15449
 ; Sequence 15449, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianteng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15449
 ; LENGTH: 291
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15449

Query Match 47.4%; Score 715; DB 15; Length 291;
 Best Local Similarity 53.2%; Pred. No. 1.9e-67;
 Matches 151; Conservative 35; Mismatches 94; Indels 4; Gaps 2;
 QY 14 RVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGPLPENVLDPACA 73
 DB 12 RTERQLDACLPSFALAPQRLHMAHVAVLGGKRVPLLVYASGALFGAEGSLDPAVA 71
 QY 74 VEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGIT 133
 DB 72 VELIHAYSLVHDDLPAMDNDLRRGKPTVHVIAFDEATAILAGDALQARAFELLABAP--- 128
 QY 134 VDAPARKMITALTTRASGSGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASVNL 193
 DB 129 ADAALRVGWLQSLASAGAGMGCGQALDIDATGQVQALQALQRMHAKTGLIRAVRM 188
 QY 194 AALSCKPDLDTCVAKKLDHYAKCIGLSFQVKDDILDIEADTATLTKTQKDIINDKPT 253
 DB 189 GVLT-GGAGIADQQLDAPADALGAFQVVDLIDVSSSAQLGKTAGDKAQAQSKSTYPA 247
 QY 254 LIGNAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 297
 DB 248 LIGMDGAKTKLAELASHMHALLPYGESANTLASLORFAVDRAH 291

RESULT 10
 US-10-369-493-15817
 ; Sequence 15817, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 15817
LENGTH: 284

TYPE: PRT
ORGANISM: Xanthomonas campestris
S-10-369-493-15817

Query Match 46.8%; Score 707; DB 15; Length 284;
Best Local Similarity 53.2%; Pred. No. 1.3e-66;
Matches 150; Conservative 35; Mismatches 93; Indels 4; Gaps 2;

Y 14 RVERALDARLPAENILPQTHQAMRVSVLNGKRTPLTYATGQALGLPENVLDPACA 73
b 7 RTERQLDACLPSPALAPQRLHAAMRHAIVLGGKRMRLPLVYASGALFGAEGSLDAPAVA 66
Y 74 VEFHVSILHDDLPAMDNDLRRGKPTCHKAYDRAATAILAGDALQALAFVLANDPGIT 133
b 67 VELIHAYSLVHDDLPAMDNDLRRGHTVTHIAFDEATAILAGDALQALAFVLANDPGIT 123
Y 134 VDAPARLKMITALTRASGQGMVGGQAIIDLSVGRKLTLPLENNHHTKGTALIRASVNL 193
b 124 ADAALRVGLQSLASAGAAGMCGGQALDIDATGQVQALEALQRMHALKTGALIRAAVRM 183
Y 194 AALSKPDLCTVAKKLDHYAKCIGLSFQVKDDILDLEADTATLTKTGKQKIDNDKPTTYP 253
b 184 GVLT-GGAGIADQORLDADFADALGLAFQVRDDILDVSSAQLGKTAGKDAQAQSKSTYPA 242
Y 254 LLGMAGAKOKAQELHQAESVLTGFGSEADLLRELSTYIER 295
b 243 LLGMDGAKTKLAELASHMHALLLPYGESGATLASIGRFAVDR 284

RESULT 11
S-10-369-493-16200
Sequence 16200, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 16200
LENGTH: 284
TYPE: PRT
ORGANISM: Xanthomonas campestris
S-10-369-493-16200

Query Match 46.8%; Score 707; DB 15; Length 284;
Best Local Similarity 53.2%; Pred. No. 1.3e-66;
Matches 150; Conservative 35; Mismatches 93; Indels 4; Gaps 2;

Y 14 RVERALDARLPAENILPQTHQAMRVSVLNGKRTPLTYATGQALGLPENVLDPACA 73
b 7 RTERQLDACLPSPALAPQRLHAAMRHAIVLGGKRMRLPLVYASGALFGAEGSLDAPAVA 66
Y 74 VEFHVSILHDDLPAMDNDLRRGKPTCHKAYDRAATAILAGDALQALAFVLANDPGIT 133

Db 67 VELIHAYSLVHDDLPAMDNDLRRGHTVTHIAFDEATAILAGDALQALAFVLANDPGIT 123
QY 134 VDAPARLKMITALTRASGQGMVGGQAIIDLSVGRKLTLPLENNHHTKGTALIRASVNL 193
Db 124 ADAALRVGLQSLASAGAAGMCGGQALDIDATGQVQALEALQRMHALKTGALIRAAVRM 183
QY 194 AALSKPDLCTVAKKLDHYAKCIGLSFQVKDDILDLEADTATLTKTGKQKIDNDKPTTYP 253
Db 184 GVLT-GGAGIADQORLDADFADALGLAFQVRDDILDVSSAQLGKTAGKDAQAQSKSTYPA 242
QY 254 LLGMAGAKOKAQELHQAESVLTGFGSEADLLRELSTYIER 295
Db 243 LLGMDGAKTKLAELASHMHALLLPYGESGATLASIGRFAVDR 284

RESULT 12
US-09-815-242-14084
Sequence 14084, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Olesen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14084
LENGTH: 299
TYPE: PRT
ORGANISM: Salmonella typhi
US-09-815-242-14084

Query Match 46.7%; Score 705.5; DB 9; Length 299;
Best Local Similarity 52.7%; Pred. No. 2e-66;
Matches 156; Conservative 39; Mismatches 94; Indels 7; Gaps 3;

QY 3 KLKAYLTVCOERVERALDARLPAENILPQTHQAMRVSVLNGKRTPLTYATGQALGL 62
Db 6 QLCQCVTQANQALSREPI-APLPQN---TPVVEAMQYCALLGCKRLRPFLVYATGQMGV 61
QY 63 PENYLDAPACAVEPIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDRAATAILAGDALQALA 122
Db 62 STATLDAPAAAVECIHAYSLIHDDLPAMDNDLRRGLPTCHIKFGEANAILAGDALQTLA 121
QY 123 FEVLNDPGLTVDAPARLKMITALTRASGQGMVGGQAIIDLSVGRKLTLPLENNHHTK 182
Db 122 FTILSDAPMEVADRDRKIAMIASLAWASGIAGCCGQALDLAEGQRIITLDALERTHEK 181
QY 183 TGAIRASVNLAAALSKPDLEPTCVAKKLDHYAKCIGLSFQVKDDILDLEADTATLTKTGK 242

```
Db 182 TGAATRAAVRLGALSAGDKGNTLPILDRYAESIGLAFQVQDDILDVVGDTATLGKQGA 241
QY 243 DIDNDKPTYPALLGMAGAKQAQELHEQAVESLTFGSEA---DLLRELSLYIER 295
Db 242 DQQLCKSTYPALLGLEQARNKARDLIEDARQSLHQLAAQSLDTSALEALADYIIOR 297

RESULT 13
US-09-815-242-10069
; Sequence 10069, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karli L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10069
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10069

Query Match 45.8%; Score 691.5; DB 9; Length 299;
Best Local Similarity 52.0%; Pred. No. 6.2e-65;
Matches 153; Conservative 36; Mismatches 96; Indels 9; Gaps 3;

QY 8 LTVQCVERALD---ARLPAENILPQTLHQAMRYSVLNGGKRTPLLTATGQALGLPE 64
Db 7 LEACVKQAQALSRIAPLPFON---TPVETMVGALGGKRLRPLVATGHRMFGVST 63
QY 65 NVLDAPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 124
Db 64 NTLDPAAAVECIHAYSLIHDDLPAMDDDLRRGLPTCHKVFGGANAILAGDALQTLAFS 123
QY 125 VLANDPGITVDAPARKMITALTTRASGQGMVGQAIDLGSGVGRKLTLPENNMHIKGTG 184
Db 124 ILSDADMEVSDRDRISMSIASASGIAGMCGQALDLDAGKHVPDLALERHHRKGTG 183
QY 185 ALIRASVNLAKSPDLDTCAVKLDHYAKCIGLSFOVKDDILDIEDATATLGKTOGKDI 244
Db 184 ALIRAAVRLGALSAGDKGRALPVLIDKYAESIGLAFQVQDDILDVVGDTATLGKQCADQ 243
QY 245 DNDKPTYPALLGMAGAKQAQELHEQAVESLTFGSEA---DLLRELSLYIER 295
Db 244 QLCKSTYPALLGLEQARKKARDLIDDAQSLKQLAQSLDTSALEALADYIIOR 297

RESULT 14
US-10-369-493-732
; Sequence 732, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 732
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-732

Query Match 45.8%; Score 691.5; DB 15; Length 299;
Best Local Similarity 52.0%; Pred. No. 5.2e-65;
Matches 153; Conservative 36; Mismatches 96; Indels 9; Gaps 3;

QY 8 LTVQCVERALD---ARLPAENILPQTLHQAMRYSVLNGGKRTPLLTATGQALGLPE 64
Db 7 LEACVKQAQALSRIAPLPFON---TPVETMVGALGGKRLRPLVATGHRMFGVST 63
QY 65 NVLDAPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 124
Db 64 NTLDPAAAVECIHAYSLIHDDLPAMDDDLRRGLPTCHKVFGGANAILAGDALQTLAFS 123
QY 125 VLANDPGITVDAPARKMITALTTRASGQGMVGQAIDLGSGVGRKLTLPENNMHIKGTG 184
Db 124 ILSDADMEVSDRDRISMSIASASGIAGMCGQALDLDAGKHVPDLALERHHRKGTG 183
QY 185 ALIRASVNLAKSPDLDTCAVKLDHYAKCIGLSFOVKDDILDIEDATATLGKTOGKDI 244
Db 184 ALIRAAVRLGALSAGDKGRALPVLIDKYAESIGLAFQVQDDILDVVGDTATLGKQCADQ 243
QY 245 DNDKPTYPALLGMAGAKQAQELHEQAVESLTFGSEA---DLLRELSLYIER 295
Db 244 QLCKSTYPALLGLEQARKKARDLIDDAQSLKQLAQSLDTSALEALADYIIOR 297

RESULT 15
US-10-369-493-10187
; Sequence 10187, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10187
; LENGTH: 306
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-10187
```

Query Match 45.7%; Score 690; DB 15; Length 306;
Best local Similarity 52.2%; Pred. No. 9.3e-65;
Matches 157; Conservative 36; Mismatches 94; Indels 14; Gaps 5;

3 KUKAYLVCQERVERALDARLPAENILPQTLHQAMRYSVLNGGRTRPLLTATGQALGL 62
14 ELKPYLADRKYLVSALDRLFPADREPRALTGMARYSLIGGRPLRPLTLACCEAVG- 72
63 PENVLDAP---ACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQ 119
73 --GALERAMNFACAMECVHTYSLIHDDLPAMDNDLRRGHPTCHRYQGEATAILAGDALL 130
120 ALAFEVLANDPGITVDAPARLKMITALTRASSQGWGQAIIDLGSVGRKLTLPLENNH 179
131 THAFTE-LAAKPVTVGLPEARLELIGELAMAAAGTYGWWGQMLDIQGEKRPLELVQLQNIH 189
180 IHKTGALIRASVNLAL---SKPDLDTCVAKKLDHYKCGISFQVKDDILDIEADTATL 236
190 IHKTGALIRVACLAGARLGGTTPD---QVHLKRYGEALGLAFQITDILDEVGDTKTL 245
237 GKTQKGDINDKPTYPALIGMAGAKQAKELHEQAVESLTGFGSEADLLRELSLYIIERT 296
246 GRATGVDRALDKATYPKLNGLAQAREARENLMQEATRCLTETFGQADPLRALAQYVIART 305
297 H 297
306 H 306

Search completed: February 29, 2004, 15:27:46
Job time : 22.7191 secs

result No.	Score	Query Match	Length	DB	ID	Description
1	1510	100.0	297	4	US-09-934-903-14	Sequence 14, Appl
2	831	55.0	401	4	US-09-250-953A-26387	Sequence 26387, A
3	702.5	46.5	307	4	US-09-543-651A-5908	Sequence 5908, Ap
4	693.5	45.9	399	4	US-09-468-039A-12733	Sequence 12733, A
5	605.5	40.1	308	4	US-09-328-352-4190	Sequence 4190, Ap
6	588.5	39.0	393	3	US-09-187-050-2	Sequence 2, Appli
7	588.5	39.0	393	3	US-09-187-050-14	Sequence 14, Appl
8	588.5	39.0	393	3	US-09-187-050-16	Sequence 16, Appl
9	588.5	39.0	393	3	US-09-187-050-18	Sequence 18, Appl
10	588.5	39.0	393	3	US-09-187-050-20	Sequence 20, Appl
11	588.5	39.0	393	3	US-09-187-050-22	Sequence 22, Appl
12	588.5	39.0	393	3	US-09-187-050-24	Sequence 24, Appl
13	588.5	39.0	393	3	US-09-187-050-26	Sequence 26, Appl
14	588.5	39.0	393	3	US-09-187-050-27	Sequence 27, Appl
15	588.5	39.0	393	3	US-09-187-050-28	Sequence 28, Appl
16	588.5	39.0	393	3	US-09-187-050-29	Sequence 29, Appl
17	588.5	39.0	393	3	US-09-187-050-30	Sequence 30, Appl
18	588.5	39.0	393	3	US-09-187-050-31	Sequence 31, Appl
19	588.5	39.0	393	3	US-09-187-050-32	Sequence 32, Appl
20	588.5	39.0	393	3	US-09-187-050-33	Sequence 33, Appl
21	588.5	39.0	393	3	US-09-187-050-34	Sequence 34, Appl
22	587	38.9	285	3	US-09-187-050-12	Sequence 12, Appl
23	578	38.3	297	1	US-08-534-910B-7	Sequence 7, Appli
24	577	38.2	297	3	US-08-686-466-2	Sequence 2, Appli
25	577	38.2	297	3	US-09-475-304-2	Sequence 3, Appli
26	577	38.2	297	4	US-08-101-126-3	Sequence 5, Appli
27	577	38.2	297	4	US-09-367-538A-5	

Db 241 GSDINDKPTYPALLGAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 297
|||||

RESULT 2

US-09-252-991A-26387
; Sequence 26387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26387
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26387

Query Match 55.0%; Score 831; DB 4; Length 401;
Best Local Similarity 60.3%; Pred. No. 2e-77;
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;
QY 4 LKAYLTVCQVERALDARLPAENILPOTLHOAMRYSVNGKTRPILLTVATQALGLP 63
DB 108 IAAVQARQARVADALDALFVAPRELQELYEAMRYSVNGKTRPILLTVATQALGLP 167
QY 64 ENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAF 123
DB 168 PORADAAACAVELIHAYSLVHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAF 227
QY 124 EYLANDPGITVDAPARLKMITALTRASSGQMGVGOALDLSGVGRKLTLPLENNHHT 183
DB 228 EVIADTRNPQHAVCLENLRLARAGSAGMGVGOALDLSGVGRKLTLPLENNHHT 287
QY 184 GALIRASVNLAAKSPDLDTCAKGLDHYAKICGLSFQVKDDILDIEADTATLGKTKGKD 243
DB 288 GALIEASVRLGALASGRAPASLAALRYAEALGAFQVQDDILDVSDTATLGKTKGKD 347
QY 244 INDKPTYPALLGAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 295
DB 348 QAHNKPTYPALLGAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 399
|||||

RESULT 3

US-09-543-681A-5908
; Sequence 5908, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5908
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5908

Query Match 46.5%; Score 702.5; DB 4; Length 307;
Best Local Similarity 52.4%; Pred. No. 2.5e-64;

Matches 150; Conservative 36; Mismatches 97; Indels 3; Gaps 1;
QY 13 ESEVERALDARLPAENILPOTLHOAMRYSVNGKTRPILLTVATQALGLPENVLDPAC 72
DB 20 QRVDEALNQALQTLFSDMPLSQAMRYGALLGGKRLRFLVAVGEMENVVAVNLDPAA 79
QY 73 AVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAFVLANDPGI 132
DB 80 AIECIHAYSLIHDDLPAMDNDLRRGKPTCHIEFGANAILAGDALQTLAIEILAKNAMP 139
QY 133 TYDAPARLKMITALTRASSGQMGVGOALDLSGVGRKLTLPLENNHHTKALIRASVN 192
DB 140 DVIAIDRVAMTAEATASGLAGMCGQALDDDBKSIDLVALEKIHLKTKALIRAAVR 199
QY 193 LAALSKPDLDTCAKGLDHYAKICGLSFQVKDDILDIEADTATLGKTKGKDINDKPTYP 252
DB 200 LGALSAGQKQHDVLPALDKYAHISGLAFQVQDDILDVSTGTEGKQCSQDEACKSTYP 259
QY 253 ALLGMAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 295
DB 260 ALLGLAQKQKAEHQAVESITGFGSEADLLRELSLYIERTH 305
|||||

RESULT 4

US-09-489-039A-12733
; Sequence 12733, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12733
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12733

Query Match 45.9%; Score 693.5; DB 4; Length 299;
Best Local Similarity 51.2%; Pred. No. 2e-63;
Matches 149; Conservative 38; Mismatches 101; Indels 3; Gaps 1;

QY 8 LTVCOERVERALDARLPAENILPOTLHOAMRYSVNGKTRPILLTVATQALGLPENVL 67
DB 7 LQACVEQANEALRRFIAFPQNTLVEAMHYGALLGGKRLRFLVATGEMFVCRITL 66
QY 68 DAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAFVLA 127
DB 67 DAPAAVECIHAYSLIHDDLPAMDNDLRRGKPTCHIKFGEANAILAGDALQTLAFSILS 126
QY 128 NDPGITVDAPARLKMITALTRASSGQMGVGOALDLSGVGRKLTLPLENNHHTKALI 187
DB 127 DAPMVDVDPDRLEAMVSLSAQAGVAGMCGQALDLSGVGRKLTLPLENNHHTKALI 186
QY 188 RASVNLAAKSPDLDTCAKGLDHYAKICGLSFQVKDDILDIEADTATLGKTKGKDIND 247
DB 187 RAAVENGALSAGERCAALPALDRYAENIGLAFQVQDDILDVSDTATLGKTKGKDQOLG 246
QY 248 KPTYPALLGAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 295
DB 247 KSTYPALLGAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 297
|||||

RESULT 5

US-09-328-352-4190
; Sequence 4190, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/326,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4190
LENGTH: 308
TYPE: PRT
ORGANISM: Acinetobacter baumannii
S-09-328-352-4190

Query Match 40.1%; Score 605.5; DB 4; Length 308;
Best Local Similarity 47.0%; Pred. No. 2.6e-54;
Matches 143; Conservative 50; Mismatches 100; Indels 11; Gaps 6;
Y 1 MSKLKA-YLTVCQERVERALDAPENILPOTLHOAMRYSVLNGGKRRPRLTYATGQA 59
b 9 LSQVNDVLKQQRICQDLTTLAFFSI-PEELKSAVHEAVNLGGKRVPCALYATASL 67
Y 60 LGLPENVLDA--PACAVEFHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDA 117
b 68 QDNP-NFAAARAAVAVELIHCYSLAHDLPAMDNDLRRGKPTCHKAYDEATAILAGDI 126
Y 118 LQALAEVLA---NDPGITVDAPARKMITALTALTRASGSGQXVGGQADLGSVGRKLTP 173
b 127 LQMAEVLGSLFDDGGQGTDAIVLKIQIL--ATASKVCGVLDLQABAKIQSD 184
Y 174 ELENHHTKGTALIRASVNLAAALSKEDLDTCAVAKLDHYAKCIGLSFQVKDDILDIEADT 233
b 185 ELENHHTKGTALIRASVNLAAALSKEDLDTCAVAKLDHYAKCIGLSFQVKDDILDIEADT 244
Y 234 ATLGKTCQKIDNDKTYTALLGMAGAKQAKQELHQAQVESLTGSGSEADLLRELSLYII 293
b 245 ETLGKTAGKDEQVQKSTYPALMGLEQAIYAKELHQAQVESLTGSGSEADLLRELSLYII 304
Y 294 ERTH 297
b 305 ARTN 308

ESULT 6
S-09-187-050-2
Sequence 2, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Hefner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 2
LENGTH: 393
TYPE: PRT
ORGANISM: Taxus canadensis
S-09-187-050-2

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
Y 15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLTYATGQALGPENVLDPACAV 74
b 112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRRVPCALCIACELVGGSQDLAMPACAM 169
Y 75 EFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 134
b 170 EMHTWSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 229

QY 135 DAPARKMITALTALTRASGSGQXVGGQADLGSVGRKLTPLENNHHTKGTALIRASVNL 193
DB 230 PSDRTLREVISLGGKTTIGSQGLGVGVVDITSEGDANVDLKTLEWHIHKTAIVLLECSVVS 289
QY 194 AALSXPDLDTCAVAKLDHYAKCIGLSFQVKDDILDIEADTATLGKTCQKIDNDKTYPA 253
DB 290 GGLGGATEDEIA-RIRRYARCVCGLLFQVVDILDVTKSSEELGKTAGKDLTDRKATYPK 348
QY 254 LLGMAGAKQAKQELHQAQVESLTGSGSEADLLRELSLYIIER 295
DB 349 LMGLEKAKFAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 7
US-09-187-050-14
Sequence 14, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Hefner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 14
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-14

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
QY 15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLTYATGQALGPENVLDPACAV 74
DB 112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRRVPCALCIACELVGGSQDLAMPACAM 169
QY 75 EFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 134
DB 170 EMHTWSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 229
QY 135 DAPARKMITALTALTRASGSGQXVGGQADLGSVGRKLTPLENNHHTKGTALIRASVNL 193
DB 230 PSDRTLREVISLGGKTTIGSQGLGVGVVDITSEGDANVDLKTLEWHIHKTAIVLLECSVVS 289
QY 194 AALSXPDLDTCAVAKLDHYAKCIGLSFQVKDDILDIEADTATLGKTCQKIDNDKTYPA 253
DB 290 GGLGGATEDEIA-RIRRYARCVCGLLFQVVDILDVTKSSEELGKTAGKDLTDRKATYPK 348
QY 254 LLGMAGAKQAKQELHQAQVESLTGSGSEADLLRELSLYIIER 295
DB 349 LMGLEKAKFAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 8
US-09-187-050-16
Sequence 16, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B

APPLICANT: Hefner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
TITLE OF INVENTION: Synthese, And Methods of Use
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 16
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)...(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-16

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLITYATQALGLPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKIHESMRYSLLAGKRVFPALCIAACELVGGSDLAMPACAM 169
75 EPIHVSILIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134
170 EMHTMSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229
135 DAPARLKMITALTTRASSQGMVGGQALDLSVG-RKLTPELENMHIHKTGALIRASVNL 193
230 PSDRTLVRVISELKGITIGSGLVGGQVVDITSEGDANVDLKTLEWHIHKTAIVLECSVWS 289
194 AALSKPDLDTVCVKKLDHYAKICIGLSFQVXDDILDIEADTATIGKTKGKIDNDKPTYP 253
290 GGILGGATEDEIA-RIRRYARCVGLLPQVVDLIDVTKSSEELGKTAGKDLTDRKATYPK 348
254 LLEGAGAKQKAOELHEQAVESLTGFGS-EADLLRELSLYIIER 295
349 LMGLEKAKEFAELATRAKEBELSSFDQIKAAPLIGLADYIAFR 391

RESULT 9
US-09-187-050-18
Sequence 18, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Hefner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
TITLE OF INVENTION: Synthese, And Methods of Use
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 18
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)...(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant

US-09-187-050-18

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLITYATQALGLPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKIHESMRYSLLAGKRVFPALCIAACELVGGSDLAMPACAM 169
75 EPIHVSILIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134
170 EMHTMSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229
135 DAPARLKMITALTTRASSQGMVGGQALDLSVG-RKLTPELENMHIHKTGALIRASVNL 193
230 PSDRTLVRVISELKGITIGSGLVGGQVVDITSEGDANVDLKTLEWHIHKTAIVLECSVWS 289
194 AALSKPDLDTVCVKKLDHYAKICIGLSFQVXDDILDIEADTATIGKTKGKIDNDKPTYP 253
290 GGILGGATEDEIA-RIRRYARCVGLLPQVVDLIDVTKSSEELGKTAGKDLTDRKATYPK 348
254 LLEGAGAKQKAOELHEQAVESLTGFGS-EADLLRELSLYIIER 295
349 LMGLEKAKEFAELATRAKEBELSSFDQIKAAPLIGLADYIAFR 391

RESULT 10

US-09-187-050-20
Sequence 20, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Hefner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
TITLE OF INVENTION: Synthese, And Methods of Use
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 20
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)...(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-20

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLITYATQALGLPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKIHESMRYSLLAGKRVFPALCIAACELVGGSDLAMPACAM 169
75 EPIHVSILIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134
170 EMHTMSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229
135 DAPARLKMITALTTRASSQGMVGGQALDLSVG-RKLTPELENMHIHKTGALIRASVNL 193
230 PSDRTLVRVISELKGITIGSGLVGGQVVDITSEGDANVDLKTLEWHIHKTAIVLECSVWS 289
194 AALSKPDLDTVCVKKLDHYAKICIGLSFQVXDDILDIEADTATIGKTKGKIDNDKPTYP 253


```
290 GGILGATEDEIA-RIRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTKATYPK 348
254 LIGMAGAKQAKQELHEQAVESITGFGS-EADLLRELSYIIER 295
349 LMGLEKAEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 11
3-09-187-050-22
Sequence 22, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heffner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-22

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSLVNGGKTRPPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVKVPALCIAACELVGGSDLPMTACAM 169
75 EFHYVSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
170 EMHTWSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
135 DAPARKMITALTTRASGSGQMGVGGQAILDLSVG-RKLTLPLENNHIHKTGALLIRASVNL 193
230 PSDRTLRLVISELGTIGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKIDNDKPTYP 253
290 GGILGATEDEIA-RIRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTKATYPK 348
254 LIGMAGAKQAKQELHEQAVESITGFGS-EADLLRELSYIIER 295
349 LMGLEKAEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 12
3-09-187-050-22
Sequence 24, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heffner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSLVNGGKTRPPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVKVPALCIAACELVGGSDLPMTACAM 169
75 EFHYVSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
170 EMHTWSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
135 DAPARKMITALTTRASGSGQMGVGGQAILDLSVG-RKLTLPLENNHIHKTGALLIRASVNL 193
230 PSDRTLRLVISELGTIGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKIDNDKPTYP 253
290 GGILGATEDEIA-RIRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTKATYPK 348
254 LIGMAGAKQAKQELHEQAVESITGFGS-EADLLRELSYIIER 295
349 LMGLEKAEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 13
US-09-187-050-26
Sequence 26, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heffner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-26

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSLVNGGKTRPPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVKVPALCIAACELVGGSDLPMTACAM 169
75 EFHYVSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
170 EMHTWSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
135 DAPARKMITALTTRASGSGQMGVGGQAILDLSVG-RKLTLPLENNHIHKTGALLIRASVNL 193
230 PSDRTLRLVISELGTIGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKIDNDKPTYP 253
290 GGILGATEDEIA-RIRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTKATYPK 348
254 LIGMAGAKQAKQELHEQAVESITGFGS-EADLLRELSYIIER 295
349 LMGLEKAEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391
```

```
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-24

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSLVNGGKTRPPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVKVPALCIAACELVGGSDLPMTACAM 169
75 EFHYVSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
170 EMHTWSLIHDDLPAMDNDLRRGKPTCKKAYDEATAILAGDALQALAFVLANDPGITY 134
135 DAPARKMITALTTRASGSGQMGVGGQAILDLSVG-RKLTLPLENNHIHKTGALLIRASVNL 193
230 PSDRTLRLVISELGTIGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKIDNDKPTYP 253
290 GGILGATEDEIA-RIRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTKATYPK 348
254 LIGMAGAKQAKQELHEQAVESITGFGS-EADLLRELSYIIER 295
349 LMGLEKAEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 13
US-09-187-050-26
Sequence 26, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heffner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-26

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSLVNGGKTRPPLITYATGQALGPENVLDPACAV 74
```

Db 112 VDAALDKAIPLE--YPEKIHESMRYSLLAGGKVRPALCIAACELVGGSDLPAMPTACAM 169
QY 75 EFHIVYSLIHDDLPAKDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134
Db 170 EMHTMSLIHDDLPCMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229
QY 135 DAPARKMTALTTRASGGQGVGGQAI DLGVSU--RKLTPLENNHHTKALIRASVNL 193
Db 230 PSDRTLRLVISELGTIGSQGLVGGQVVDITSEGDANVDLKTLEWHHTKAVLECSVVS 289
QY 194 AALSPPDLDTCAVAKLDHYAKICIGLSFQVDDILDIADTATLGKTDNDKPTTYP 253
Db 290 GGIIGGATEDEIA--RIRRYARCVGLLPQVDDILDVTSSSELGKTACKDILTATKATY 348
QY 254 LLMGAKAKOQAEHLEHQAQVESLTGFGS--EADLLRELSLYIER 295
Db 349 LMGLEKAKEFAAEALATRAKEELSSPDQIKAAAPLLGLADYIAFR 391

RESULT 14

US-09-187-050-27
; Sequence 27, Application US/09187050B
; Patent No. 6043072
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
; FILE REFERENCE: WSUR12423
; CURRENT APPLICATION NUMBER: US/09/187,050B
; CURRENT FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 27
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
; OTHER INFORMATION: synthase protein variant
; NAME/KEY: VARIANT
; LOCATION: (1)..(393)
; OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
; OTHER INFORMATION: variant
US-09-187-050-27

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
QY 15 VERALDARLPAENILPOTLHOAMRYSVLNGGKRTPEPLTYATGQALGPLPENVDAPACAV 74
Db 112 VDAALDKAIPLE--YPEKIHESMRYSLLAGGKVRPALCIAACELVGGSDLPAMPTACAM 169
QY 75 EFHIVYSLIHDDLPAKDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134
Db 170 EMHTMSLIHDDLPCMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229
QY 135 DAPARKMTALTTRASGGQGVGGQAI DLGVSU--RKLTPLENNHHTKALIRASVNL 193
Db 230 PSDRTLRLVISELGTIGSQGLVGGQVVDITSEGDANVDLKTLEWHHTKAVLECSVVS 289
QY 194 AALSPPDLDTCAVAKLDHYAKICIGLSFQVDDILDIADTATLGKTDNDKPTTYP 253
Db 290 GGIIGGATEDEIA--RIRRYARCVGLLPQVDDILDVTSSSELGKTACKDILTATKATY 348
QY 254 LLMGAKAKOQAEHLEHQAQVESLTGFGS--EADLLRELSLYIER 295
Db 349 LMGLEKAKEFAAEALATRAKEELSSPDQIKAAAPLLGLADYIAFR 391

RESULT 15
US-09-187-050-28
; Sequence 28, Application US/09187050B
; Patent No. 6043072
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
; FILE REFERENCE: WSUR12423
; CURRENT APPLICATION NUMBER: US/09/187,050B
; CURRENT FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 28
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
; OTHER INFORMATION: synthase protein variant
; NAME/KEY: VARIANT
; LOCATION: (1)..(393)
; OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
; OTHER INFORMATION: variant
US-09-187-050-28

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;
QY 15 VERALDARLPAENILPOTLHOAMRYSVLNGGKRTPEPLTYATGQALGPLPENVDAPACAV 74
Db 112 VDAALDKAIPLE--YPEKIHESMRYSLLAGGKVRPALCIAACELVGGSDLPAMPTACAM 169
QY 75 EFHIVYSLIHDDLPAKDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134
Db 170 EMHTMSLIHDDLPCMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229
QY 135 DAPARKMTALTTRASGGQGVGGQAI DLGVSU--RKLTPLENNHHTKALIRASVNL 193
Db 230 PSDRTLRLVISELGTIGSQGLVGGQVVDITSEGDANVDLKTLEWHHTKAVLECSVVS 289
QY 194 AALSPPDLDTCAVAKLDHYAKICIGLSFQVDDILDIADTATLGKTDNDKPTTYP 253
Db 290 GGIIGGATEDEIA--RIRRYARCVGLLPQVDDILDVTSSSELGKTACKDILTATKATY 348
QY 254 LLMGAKAKOQAEHLEHQAQVESLTGFGS--EADLLRELSLYIER 295
Db 349 LMGLEKAKEFAAEALATRAKEELSSPDQIKAAAPLLGLADYIAFR 391

Search completed: February 29, 2004, 14:54:48

Job time : 11.0253 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:26:38 ; Search time 35.1152 Seconds
(Without alignments)
2389.754 Million cell updates/sec

title: US-09-941-947A-20

effect score: 1510

sequence: 1 MSLKXAYLTVCQVERALD.....FGSEADLLRLSLVIERTH 297

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB	ID	Description
1	1510	100.0	297	5	AAE22308	Methylomo
2	1510	100.0	297	5	ABG61586	High Grow
3	1510	100.0	297	5	AAU80331	Methylomo
4	831	55.0	295	4	AAU36378	Pseudomon
5	831	55.0	295	6	ABU38665	Protein e
6	811	53.7	295	6	ABU40141	Protein e
7	791	52.4	295	6	ABU41505	Protein e
8	733	48.5	298	6	ABP78160	N. gonorr
9	733	48.5	298	6	ABP80110	N. gonorr
10	733	48.5	298	6	ABU37216	Protein e
11	729	48.3	298	6	ABU38181	Protein e
12	719.5	47.6	293	6	ABU21390	Protein e
13	711	47.1	294	6	ABU19741	Protein e
14	711	47.1	299	6	ABU28056	Protein e
15	705.5	46.7	299	4	AAU38491	Salmonell
16	705.5	46.7	299	6	ABU47496	Protein e
17	705.5	46.7	310	6	ABM67133	Photorhab
18	704	46.6	234	6	ABU49248	Protein e
19	702.5	46.5	306	6	ABU50231	Protein e
20	702.5	46.5	307	6	ABU40769	Protein e
21	699	46.3	294	6	ABU22567	Protein e
22	693.5	45.9	299	6	ABU31979	Protein e
23	691.5	45.8	299	2	AAV52835	Escherich
24	691.5	45.8	299	4	AAU34476	E. coli c
25	691.5	45.8	299	5	ABU05280	Prenyl al

ALIGNMENTS

RESULT 1

AAE22308

ID AAE22308 standard; protein; 297 AA.

XX AC AAE22308;

XX XX

DT 07-AUG-2003 (revised)

DT 25-JUL-2002 (first entry)

XX XX

DE Methylomonas 16a sp. farnesy diphosphate synthetase (fspa) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; farnesyl diphosphate synthetase; fspa.

XX OS

XX Methylomonas sp.

XX WO200218617-A2.

XX PD

XX 07-MAR-2002.

XX XX

PF 04-SEP-2001; 2001WO-US027420.

XX XX

PR 01-SEP-2000; 2000US-0229858P.

XX 01-SEP-2000; 2000US-0229907P.

XX PA

XX (DUPO) DU FONT DE NEMOURS & CO S I.

XX XX

PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odum JM, Picataggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35506.

XX DR

XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX substrates.

XX PT

XX Claim 35; Page 125-126; 156pp; English.

XX PS

XX The invention relates to a method for producing carotenoid compounds. The

XX method comprises a transformed metabolising host cell, comprising

XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

XX encoding an enzyme in the carotenoid biosynthetic pathway, under the

XX control of regulatory sequences, and contacting the host cell with carbon

XX substrate to produce a carotenoid compound. The method is useful for

XX producing carotenoid compounds such as anthraxanthin and astaxanthin, by

AAO22194 Protein r
ABG68085 E.coli pr
ABU28527 Protein e
ADE86721 geranyltr
ABU23165 Protein e
ABU39094 Protein e
ADE86723 Geranyltr
ABU05283 Prenyl al
ABG68093 E.coli pr
ABU05282 Prenyl al
ABG68092 E.coli pr
ABU05281 E.coli f
ABG68091 E.coli f
AAU35646 Haemophil
ABU30591 Protein e
ABU74157 X. fastid
AAE25932 Soybean G
ABU67329 Soybean G
ADA32903 Acinetoba
AAG42844 Arabidops

26 691.5 45.8 299 5 AAO22194
27 691.5 45.8 299 5 ABG68085
28 691.5 45.8 299 6 ABU28527
29 691.5 45.8 299 7 ADE86721
30 691 45.8 303 6 ABU23165
31 690 45.7 295 6 ABU39094
32 689.5 45.7 299 7 AD866723
33 683.5 45.3 299 5 ABU05283
34 683.5 45.3 299 5 ABG68093
35 682.5 45.2 299 5 ABU05282
36 682.5 45.2 299 5 ABG68092
37 681.5 45.1 299 5 ABU05281
38 681.5 45.1 299 5 ABG68091
39 677 44.8 295 4 AAU35646
40 677 44.8 295 6 ABU30591
41 675 44.7 291 7 AD874157
42 610 40.4 369 5 AAE25932
43 610 40.4 369 6 ABU67329
44 605.5 40.1 308 6 ADA32903
45 604 40.0 295 3 AAG42844

CC using microorganism having a nucleic acid molecule encoding enzymes in
 CC the carotenoid biosynthetic pathway and which metabolise single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Methylomonas 16a sp. geranyltransferase or farnesyl diphosphate
 CC synthetase (ispA) enzyme used in the invention. (Updated on 07-AUG-2003
 CC to correct OS field.)
 XX
 SQ Sequence 297 AA;

Query Match 100.0%; Score 1510; DB 5; Length 297;
 Best Local Similarity 100.0%; Pred. No. 7.3e-140;
 Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKIKAYLVTCQERVERALDARLPAENILPQTLHOAMEYSLVNGGKTRPLLTATGQAL 60
 DB 1 MSKIKAYLVTCQERVERALDARLPAENILPQTLHOAMEYSLVNGGKTRPLLTATGQAL 60
 QY 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 120
 DB 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 120
 QY 121 LAFEVLPDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 180
 DB 121 LAFEVLPDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 180
 QY 181 HTGALIRASVNLAAALSKPDLDTCVAKLDHVAKICIGLSFQVKDILDIADTATLQKTQ 240
 DB 181 HTGALIRASVNLAAALSKPDLDTCVAKLDHVAKICIGLSFQVKDILDIADTATLQKTQ 240
 QY 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297
 DB 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 2
 ABG61586
 ID ABG61586 standard; protein; 297 AA.

XX ABG61586;
 AC
 XX
 AC
 XX
 DT 07-AUG-2003 (revised)
 DT 27-AUG-2002 (first entry)
 XX High growth methanotrophic bacterial strain polypeptide #36.
 DE
 XX High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;
 KW methane; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;
 KW pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;
 KW ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;
 KW methane-containing environment; waste water treatment system; isoprenoid;
 KW nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.

XX Methylomonas sp.
 OS
 XX
 PN WC200220728-A2.
 XX
 PD 14-MAR-2002.
 XX
 PF 28-AUG-2001; 2001WO-US026827.
 PR
 XX 01-SEP-2000; 2000US-0229858P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 PA
 XX Koffas M, Odom JM, Schenzle A;
 PI WPI; 2002-454358/48.
 XX
 DR N-PSDB; ABX83265.
 DR
 XX New high growth methanotrophic bacterial strain, useful for producing

PT single cell proteins, grows on a C1 carbon substrate, and comprises a
 PT functional gene encoding in Embden-Meyerhof carbon pathway.
 XX
 XX Claim 11; Page 147-148; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,
 CC which grows on a C1 carbon substrate e.g. methane and methanol, and
 CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a
 CC gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S
 CC RNA. The bacterial strain is useful for the production of single cell
 CC protein and for the biotransformation of a nitrogen-containing compound,
 CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the
 CC production of a feed product comprising a protein, carbohydrates and a
 CC pigment and for reducing oxygen demand, for removing nitrates and
 CC nitrates in methane-containing environments such as landfills, waste
 CC water treatment systems or anywhere that methane, oxygen and nitrates are
 CC present. The bacterial strain of the invention can be used as a
 CC denitrifying agent for the conversion of nitrate or nitrite to nitrous
 CC oxide with methane or methanol as a carbon source. It is also used in the
 CC production of biomass including proteins, carbohydrates and a wide
 CC variety of pigments (particularly for isoprenoid pigments for the purpose
 CC of generating animal feeds), in production of terpenoid and carotenoid
 CC compounds, useful as pigments and as monomers in polymeric materials
 CC in production of exopolysaccharides at high levels. Sequences ABG61551-
 CC ABG61590 represent high growth methanotrophic bacterial strain proteins
 CC of the invention. (Updated on 07-AUG-2003 to correct OS field.)
 XX

SQ Sequence 297 AA;

Query Match 100.0%; Score 1510; DB 5; Length 297;
 Best Local Similarity 100.0%; Pred. No. 7.3e-140;
 Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKIKAYLVTCQERVERALDARLPAENILPQTLHOAMEYSLVNGGKTRPLLTATGQAL 60
 DB 1 MSKIKAYLVTCQERVERALDARLPAENILPQTLHOAMEYSLVNGGKTRPLLTATGQAL 60
 QY 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 120
 DB 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 120
 QY 121 LAFEVLPDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 180
 DB 121 LAFEVLPDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATATLQAL 180
 QY 181 HTGALIRASVNLAAALSKPDLDTCVAKLDHVAKICIGLSFQVKDILDIADTATLQKTQ 240
 DB 181 HTGALIRASVNLAAALSKPDLDTCVAKLDHVAKICIGLSFQVKDILDIADTATLQKTQ 240
 QY 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297
 DB 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 3
 AAU80331
 ID AAU80331 standard; protein; 297 AA.

XX AAU80331;
 AC
 XX
 DT 15-JUL-2002 (first entry)
 XX
 XX Methylomonas 16a ORF7 ispA protein sequence.
 DE
 XX Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;
 KW keratenoid; pigment; flavour; fragrance; open reading frame 7; ORF7;
 KW ispA; geranyltransferase/farnesyl diphosphate synthase enzyme.
 XX
 OS Methylomonas sp.
 XX
 XX WC200220733-A2.
 PN
 XX 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US026852.
XX
XX
XX 01-SEP-2000; 2000US-0229907P.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Cheng Q, Koffas M, Norton KC, Odum JM, Picataggio SK;
XX Rouviere PE, Schenzle A, Tomb J;
XX WPI; 2002-383051/41.
XX N-PSDB; ABK50087.
XX Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,
XX isolated from *Methylobacterium* 16a, useful for the production of isoprenoid
XX compounds.
XX
XX Claim 4; Page 76-77; 84pp; English.
XX
XX The present invention relates to a new nucleic acid molecule encoding an
XX isoprenoid biosynthetic enzyme isolated from *Methylobacterium* 16a. The
XX invention is useful for obtaining a nucleic acid molecule encoding an
XX isoprenoid compound biosynthetic enzyme, and for the microbial production
XX of isoprenoid compounds. The molecules of the invention are also useful
XX for regulating isoprenoid biosynthesis in an organism and for producing
XX recombinant organisms for producing various isoprenoid compounds. The
XX nucleic acid is also useful for feed additive, for the production of
XX xeratenoids and their derivatives, isoprenoid intermediates, and as pure
XX products useful as pigments, flavours and fragrances. The present amino
XX acid sequence represents the *Methylobacterium* 16a open reading frame 7 (ORF7)
XX isPA (geranyltransferase or farnesyl diphosphate synthase enzyme) protein
XX of the invention, as described above
XX
XX Sequence 297 AA;
XX
XX Query Match 100.0%; Score 1510; DB 5; Length 297;
XX Best Local Similarity 100.0%; Pred. No. 7.3e-140;
XX Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2Y 1 MSKUKAYITVCOEVRERALDARLPAENILPOTLHOAMRYSVLNGSKETRELLTYATQAL 60
2b 1 MSKUKAYITVCOEVRERALDARLPAENILPOTLHOAMRYSVLNGSKETRELLTYATQAL 60
2Y 61 GLPENVLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQA 120
2b 61 GLPENVLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQA 120
2Y 121 LAFEVLNDPQITVDAPARLKMITALTFRASQSGVMGQQAIDLGSGVRKLTIPLENNHI 180
2b 121 LAFEVLNDPQITVDAPARLKMITALTFRASQSGVMGQQAIDLGSGVRKLTIPLENNHI 180
2Y 181 HKTGALIRASVNLAAALSKPDLDTCAKLDHYAKICIGLSFQVKKDILDIADTATLGKTK 240
2b 181 HKTGALIRASVNLAAALSKPDLDTCAKLDHYAKICIGLSFQVKKDILDIADTATLGKTK 240
2Y 241 GKDIDNDKPTYPALLGMAGAKQKAEHQVSLATGFGSEADLLRELSLYIERTH 297
2b 241 GKDIDNDKPTYPALLGMAGAKQKAEHQVSLATGFGSEADLLRELSLYIERTH 297
RESULT 4
AAU36378
X AU36378 standard; protein; 295 AA.
X
X AAU36378;
X
X 14-FEB-2002 (first entry)
X
X Pseudomonas aeruginosa cellular proliferation protein #368.
X
X Antisense; prokaryotic cellular proliferation protein; antibiotic;
X antibacterial; drug design.
X

OS Pseudomonas aeruginosa.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009180.
XX
XX 21-MAR-2000; 2000US-0191078P.
XX 23-MAY-2000; 2000US-0206848P.
XX 26-MAY-2000; 2000US-0207727P.
XX 23-OCT-2000; 2000US-0242578P.
XX 27-NOV-2000; 2000US-0253625P.
XX 22-DEC-2000; 2000US-0257931P.
XX 16-FEB-2001; 2001US-0269308P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX N-PSDB; AAS54237.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Example 3; SEQ ID NO 11971; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes,
XX themselves and the encoded proteins. The prokaryotes used are *Escherichia*
XX coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
XX *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence represents an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 295 AA;
XX
XX Query Match 55.0%; Score 831; DB 4; Length 295;
XX Best Local Similarity 60.3%; Pred. No. 4.7e-73;
XX Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;
QY 4 LKAYITVCOEVRERALDARLPAENILPOTLHOAMRYSVLNGSKETRELLTYATQALGLP 63
Db 2 IAAAYQARQARVDAALDALFVAPEELQRYEAMRYSMGKGRVPLLAYAACALGGA 61
QY 64 ENVLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQA 123
Db 62 QPADAACAVELIHAYSLVHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQA 121
QY 124 EVLANDPQITVDAPARLKMITALTFRASQSGVMGQQAIDLGSGVRKLTIPLENNHI 183
Db 122 EVLANDPQITVDAPARLKMITALTFRASQSGVMGQQAIDLGSGVRKLTIPLENNHI 181
QY 184 GALTIRASVNLAAALSKPDLDTCAKLDHYAKICIGLSFQVKKDILDIADTATLGKTK 243
Db 182 GALTIRASVNLAAALSKPDLDTCAKLDHYAKICIGLSFQVKKDILDIADTATLGKTK 241
QY 244 IDNDKPTYPALLGMAGAKQKAEHQVSLATGFGSEADLLRELSLYIERTH 295
Db 242 QAHNKPTYPALLGLEAAKVALELRDLAALDGFPPSPADPLRLQARYIVER 293

XX SQ Sequence 295 AA;
Query Match 55.0%; Score 831; DB 6; Length 295;
Best local Similarity 60.3%; Pred. No. 4.7e-75;
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;
XX AC ABU38665;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #24192.
XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Pseudomonas aeruginosa.
XX SN W0200277183-A2.
XX PD 03-OCT-2002.
XX XX 21-MAR-2002; 2002MO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX N-PSDB; ACA42535.
XX WPI; 2003-029926/02.
XX DE New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids, required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 66589; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumonise* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 295 AA;
Query Match 55.0%; Score 831; DB 6; Length 295;
Best local Similarity 60.3%; Pred. No. 4.7e-75;
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;
QY 4 LKAYLTVCQBRVERALDARLPANILPOTLHOAMRYSVLNGKRTPLITYATGQALGP 63
DB 2 IAAQARCCQARVDAALDALFVAPREELQRLYEAMRYSVNGKRVPLLAYAACALGGA 61
QY 64 ENVLDPACAVEPIHYVSLIHDDLPAKNDLIRGKPTCKHAYDEATILAGDALQALAF 123
DB 62 PQADAAACAVELIHAYSLVHDDLPAKNDLIRGKPTCKHAYDEATILAGDALQALAF 121
QY 124 EVLNDPGITVDAPARLKMITALTRASSGQGVGGQAIIDLGSGVGRKLTLPELENMHIHT 183
DB 122 EVLADTERNPQEHAVCVLEMLTLARAGSAGWGGQAIIDLGSGVGGVADQAALEVMERHT 181
QY 184 GALIRASVNLALSKPDLDTCAKLDHYAKICIGLSFQVKDILIDIEATILGKTQGD 243
DB 182 GALTASVRLGALASGRABPASLAALERYAEATGLAFQVQDDILOVESDTATLGTQGD 241
QY 244 IDNDKPTYPALLGMAKAKQKQAEHQAVESLTGFGSEADLLRELSLYIIR 295
DB 242 QAHNKPTYPALLGLEAKGVLELRDLALALDGGFPSPADPLQLARYIVER 293
RESULT 6
ABU40141
ID ABU40141 standard; protein; 295 AA.
XX AC ABU40141;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #25568.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Pseudomonas putida.
XX SN W0200277183-A2.
XX PD 03-OCT-2002.
XX XX 21-MAR-2002; 2002MO-US0009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX N-PSDB; ACA44011.
XX WPI; 2003-029926/02.
XX DE New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 68065; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumonise* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Db 182 GALIEASVRLGALSGQADQRLDALQVYARAVGLAFQVQDDILDVESDTATLGRQGD 241
QY 244 INDNKPPTVYKLMGLEAARSYAHKLVAEAVALLBPFQDKALELRQLAEFAVARKY 298
Db 242 IARDKPTYPALLGLDAAKGAYALELRDQALSLFDDTREPRLRLARYIVERRH 295
RESULT 8
ID ABP78160
XX ABP78160 standard; protein; 298 AA.
AC ABP78160;
XX
DT 07-MAR-2003 (first entry)
DE N. gonorrhoeae amino acid sequence SEQ ID 2850.
XX Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX 12-FEB-2002; 2002WO-IB002069.
XX 12-FEB-2001; 2001GB-00003424.
XX (CHIR-) CHIRON SPA.
XX Pontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; AB239130.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX Disclosure; Page 396; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX Sequence 298 AA;
Query Match 48.5%; Score 733; DB 6; Length 298;
Best Local Similarity 51.7%; Pred. No. 2.1e-63;
Matches 152; Conservative 44; Mismatches 96; Indels 2; Gaps 1;
QY 4 LKAYLTVCQERVARLDARLPARNILPOTLHQAMKYSVLNGGKTRPLLTATGQALGLP 63
Db 7 LKAWQRAQAQTELLERFLPSGNEIPTLHEAMRYAALDGGKLRPMLVLAASELGAM 66
QY 64 ENVLDPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123
Db 67 ADAGQAAVAIEMIHVSLVHDDNPAMDNDLSLRGKPTCHKIYGEATALLTGDAQTQAF 126
QY 124 EYLANDPGITVDAPARLKMITALTRASQSGQVWGQADLDSVGKRLTLPLENWHIHT 183
Db 127 DVLSP--TELPAAQLAMLSVLAKAGSAGWAGQALDLANVGKQWQADLERHSLKT 184
QY 184 GALIRASVNLAAALSKPDLDTCAVKKLDHYAKICIGLSFQVKDDILDIEADTATLTKTQGD 243
Db 185 GALIRAAVLGATACPDLSDAELAVLDAYAAKGLAFQVDDVLDCEADTATLTKTAGKD 244
QY 244 INDNKPPTVYKLMGLEAARSYAHKLVAEAVALLBPFQDKALELRQLAEFAVARKY 297

Db 245 ADNDKPTTYKLMGLEAARSYAHKLVAEAVALLBPFQDKALELRQLAEFAVARKY 298
RESULT 9
ID ABP80110
XX ABP80110 standard; protein; 298 AA.
AC ABP80110;
XX
DT 07-MAR-2003 (first entry)
DE N. gonorrhoeae amino acid sequence SEQ ID 6750.
XX Antibacterial; infection; vaccine; gene therapy.
XX Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX 12-FEB-2002; 2002WO-IB002069.
XX 12-FEB-2001; 2001GB-00003424.
XX (CHIR-) CHIRON SPA.
XX Pontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; AB241080.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX Disclosure; Page 678; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX Sequence 298 AA;
Query Match 48.5%; Score 733; DB 6; Length 298;
Best Local Similarity 51.7%; Pred. No. 2.1e-63;
Matches 152; Conservative 44; Mismatches 96; Indels 2; Gaps 1;
QY 4 LKAYLTVCQERVARLDARLPARNILPOTLHQAMKYSVLNGGKTRPLLTATGQALGLP 63
Db 7 LKAWQRAQAQTELLERFLPSGNEIPTLHEAMRYAALDGGKLRPMLVLAASELGAM 66
QY 64 ENVLDPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123
Db 67 ADAGQAAVAIEMIHVSLVHDDNPAMDNDLSLRGKPTCHKIYGEATALLTGDAQTQAF 126
QY 124 EYLANDPGITVDAPARLKMITALTRASQSGQVWGQADLDSVGKRLTLPLENWHIHT 183
Db 127 DVLSP--TELPAAQLAMLSVLAKAGSAGWAGQALDLANVGKQWQADLERHSLKT 184
QY 184 GALIRASVNLAAALSKPDLDTCAVKKLDHYAKICIGLSFQVKDDILDIEADTATLTKTQGD 243
Db 185 GALIRAAVLGATACPDLSDAELAVLDAYAAKGLAFQVDDVLDCEADTATLTKTAGKD 244
QY 244 INDNKPPTVYKLMGLEAARSYAHKLVAEAVALLBPFQDKALELRQLAEFAVARKY 297
Db 245 ADNDKPTTYKLMGLEAARSYAHKLVAEAVALLBPFQDKALELRQLAEFAVARKY 298

QY 192 NL AALSKPDLTCVAKKLDHYAKCIGLSFQVQDDILDIEADTATLGKTQKIDNDKPTY 251

Search completed: February 29, 2004, 14:43:39
Job time : 41.2152 secs
